



特集

デジタル・ スカラシップと データの 新しいかたち

私達、社会調査に携わる者にとって、デジタル技術を用いずに調査データを収集・分析することは考えられない。筆者の専門である社会学を例にとると、「デジタル社会学」は確かに社会学の下位分野として存在し、その名を冠した書物も多数出版されているが、それは決して“社会学にデジタル技術を取り入れたもの”ではない。

一方、社会科学よりも格段に長い歴史をもち、紙や羊皮紙に書かれたテキストや絵画などを扱ってきた人文学では、その伝統の深さ故にか、「デジタル・ヒューマニティーズ」というように自らの領域に「デジタル」という言葉を冠することに躊躇しない。

本特集では、人文学の分野から発信されるデジタル・スカラシップの研究成果を足がかりに、コンピュータ技術によってもたらされる従来の〈人文学〉と〈社会科学〉との対話と交配の可能性、そしてデジタル技術の社会調査への貢献と課題について検討する。

本特集を通じて、データや情報の温故知新を堪能していただければ幸いである。

(平尾桂子)



特集論文

1

デジタル・スカラシップ 越境する知識生産システム

平尾桂子

上智大学大学院地球環境学研究所 教授

1 はじめに

デジタル技術の発展は、人類がこれまで蓄積してきた「知のかたち」を根底から変えようとしている。その動きは、情報の保存・伝達媒体が従来の紙の印刷物からデジタル信号に置き換わる、単なる「デジタル化」にとどまらない。データ収集、分析手法、結果の提示、研究成果の出版や普及など、学術研究のすべてに関わる側面を変え、人文学、社会科学、芸術、自然科学の境界線を引き直し、従来の〈学問体系〉をも組み替える「知の革命」として進行している。

その一つの現れとして、世界的潮流になりつつあるデジタル技術を人文学に応用したデジタル・ヒューマニティーズ (digital humanities) が挙げられよう。デジタル・ヒューマニティーズとは、人文科学 (humanities) にデジタル技術を取り込んだ新しい学問領域である。具体的な内容は多岐にわたるが、紙媒体を中心に展開されてきた従来の人文科学の学術的伝統を礎として、コンピュータ技術によって可能になったデータの収集、分析、教育、出版などの新しい取り組みである (Burdick et al., 2012)。デジタル技術という方法論によって牽引されるため、基本的に学際的であり、分野横断的であり、その周縁は限りなく広い。

デジタル・ヒューマニティーズという言葉が登場したのは2000年以降と比較的最近のことである。しかし、人文学で最初にコンピュータを用いた研究は、イエズス会士のRoberto BusaとIBMの創始者の一人であるThomas J. Watsonが共同で作成したThomas Aquinasの膨大な著作の用語索引Index Thomisticusで、その作業の開始は実に1949年まで遡る (Burdick et al., 2012)。デジタル・ヒューマニティーズという学問領域は、言葉こそ新しいが、20世紀中ごろにおけるコンピュータ技術の登場とともに出現し、その発達とともに発展してきたといえる。

デジタル・ヒューマニティーズは「ヒューマニティーズ」(人文学) と銘打ってはいるものの、人文学に特化したものではない。事実、人文学での試みとして発達したこの潮流は、その総称である〈デジタル・スカラシップ〉としてとらえる方がふさわしい (Pearceほか, 2012)。

本特集では、人文学の分野から発信されるデジタル・スカラシップの研究成果を足がかりに、デジタル技術によって開かれた〈人文学〉と〈社会科学〉との対話と交配の可能性、そしてデジタル技術の発展がもたらす社会調査への貢献と課題について検討する。もともとはデジタル・ヒューマニティーズを特集のテーマにと考えていたが、この分野の研究対象が、厳密な意味での人文学にとどまらないため、本特集ではあえ

て広いタームの「デジタル・スカラシップ」という言葉を採用した。

従来、人文学と社会科学は研究対象や研究方法の違いにより、異なる学問分野と位置づけられてきた。人文学は人間の精神や文化を、社会科学は人間集団や社会の在り方をそれぞれ主な研究対象とし、前者は定性的な記述を用いて事実の解釈や意味を、後者は(どちらかという)と定量的な記述を通して現実を検証可能なものとして捉えようとしてきた。それぞれにデジタル技術を取り込むことで、人文学はテキストマイニングによるソーシャルネットワーク分析など、定量的・統計的手法を歴史学や文学の研究に持ち込むことに挑戦してきた。一方、社会科学でもオーラル・ヒストリーやナラティブなど質的なものを定量的に分析する研究も増えている。データ処理技術と分析手法で両者を架橋することにより、従来の社会調査で収集したデータをより重層的に分析し、より視覚的かつ重層的に発表することが可能になっている。

この特集では特に、情報科学と人文学との接合がもたらす新しい「知のかたち」を紹介するとともに、今まで調査の専門家がデータとして取り扱わなかった新しいデータとのクロスファーマイライゼーションの可能性も考えてみたい。

まずはデータマネジメントの歴史を振り返った上で、デジタル技術の発展によってもたらされたオープンアクセス・モデルと調査データの新しい提示方法の二点について概観する。

2 データと情報の温故知新

健康機器や家電機器、オーディオ製品、調理器具、交通手段など、ありとあらゆるモノと人間とがつながるIoT (Internet of Things) の登場により、私達はサイバー空間(仮想空間)とフィジカル空間(現実空間)が高度に融合された「Society 5.0」(総務省,2018)に生きるようになった。

しかし、「Society 1.0」(狩猟社会)はともかく

として、「Society 2.0」(農耕社会)以降、少なくとも文字の発明後は、人類にとって「データ」や「情報」の意味やその価値は、本質的にはさほど変わっていないように思えてならない。

たとえば、古代メソポタミアの遺跡からは、楔形文字が刻まれた、おびただしい数の粘土板が発掘されている。それらは神殿や王宮など、一定の空間に収められていたことから、そこが粘土板を「保存」して「利用」する図書館のような役割を担っていたと考えられている。

また、古代ヘレニズム文化の中心となったアレキサンドリア図書館では、プトレマイオス王朝の王たちが蔵書を拡充すべく、莫大な財産を投じてありとあらゆる主題のありとあらゆる著者の文献を収集していった。アレキサンドリア港に停泊する船舶を検分し、積荷に書物があれば没収し、原本ではなく写本を返すといった乱暴な手段に訴えることもあったという(ブラセル,1998)。

こうして集められた書物(巻物)は系統的に「分類」され、「コピー(写本)」され、「格納(アーカイブ)」され、「利用」(特権階層に限られてはいただろう)されてきた。言い換えると、人類は文字を発明するとほぼ同時に、文字が記されたモノを集め、火災などで失われることに備えてコピーを作り、後に利用できるように分類し、大切に保管するという営みを営々と続けたのである。

楔形文字のデータを記した粘土板は、記録したシュメールの商人にとってはただの会計記録だっただろう。だが、その粘土板を神殿の書庫に収納したシュメールの神官にとっては、記録としての「価値」があった。その価値は、商人の人生とは切り離され、商人の死後も存続するものと考えた。彼(多分男性だっただろう)は、自分以外の神官が粘土板を取り出して利用できるように、整理し、分類し、格納した。さらに、神官を召し抱えた王にとって、粘土板の情報は都市国家の運営に欠かせないものだったはずだ。その記録を後に王国の継承者が参照できるように、粘土板を整理して一箇所に保管するよう命



じたに違いない。こうして保管された粘土板の文字が約5000年近い時を経て再び解読され、古代都市国家に関する新たな知識をもたらすなどということ、誰も想像できなかったに違いない。

「データ」や「情報」、あるいは「情報時代」という言葉に対し、私達はとするとコンピュータによってもたらされたデジタル技術によるものを思い浮かべがちだが、データや情報にたいして現代の私たちが抱くのと同じような情念を人類はずっと抱き続けてきたのではないだろうか。人は自分の経験したことを他人に伝えるために文字を書き記し、他人が書き記したモノを集めて蓄積・分類・整理し、一定の規則に沿って並べて保管してきた。書物は世界を知る手がかりであり、それを書き記した人物や収集した人物が死んだあとも、後世に伝えられるべきものと考えられた。そして、そこから得られる新たな知識はいつか新しい知識によって上書きされることも覚悟していたに違いない。そのような「知の営み」は、データを媒介するメディアを大切に保存・分類・格納し、再利用するシステムを構築することによってこそ可能になるということも知っていたはずだ。

データマネジメントの歴史を紐解くと、人類は常に「データ・オーバーロード (情報過多)」に悩まされてきたことがわかる。ヨーロッパ中世の神学者も、ルネッサンス期に活躍したスペインの翻訳者集団も、18世紀の百科全書派の学者たちも、現代の私達と同じように「あふれる情報をいかに管理するか」という難題の解決に腐心してきた。(Blair, 2010)。現代ではコンピュータ技術の発達により、データの流通量も爆発的に増加し、データ量の規模もその多寡を問題にする領域を超えている。300エクサバイトと1000エクサバイトの差を身体感覚として把握できないように、私達は一定以上のデータ量を「たくさん」と認識し、それを「ビッグデータ」と呼んでいる。ルネッサンス期の学者たちも、イスラム文化圏からもたらされる古典ギリシャ文献の怒

涛に直面したとき、書物の実数は関係なく、それらを「たくさん」のデータ量、つまりビッグデータだと認識していたに違いない。

その「たくさん」を処理する方法として編み出されたのが書誌の分類法やそれに基づく目録の作成である。これによって、書物を検索し、識別し、所在を知り、読み直し、そこから得られた知見を後世に伝えられるようになった。その方法も時代により変化していったが、この基本的な機能は目録が紙ではなく、電子信号に置き換えられたコンピュータ時代になっても変わらない。

ここで、簡単にデジタル・ヒューマニティーズ(スカラーシップ)の歴史を振り返ってみよう¹⁾。

前述のように、人文学にデジタル技術を利用した最初の研究として知られる *Index Thomisticus* の開始は大型コンピュータの黎明期である1949年に遡る。1970年代には、言語学と文学におけるコンピュータの応用に関する学会がイギリスとアメリカで設立され、1980年代には小型コンピュータが登場し、研究者個人が利用できるインフラが整った。

1989年には、テキストデジタル化のガイドラインの策定に向けたプロジェクト TEI (Text Encoding Initiative) が発足した。それまでは、様々なタイプの文書をデジタル化するための標準規格がなく、メタデータの扱いも統一されていなかったためである。TEIはその後、テキスト資料を効果的に共有するための事実上の国際基準としての役割を担っている。

哲学、言語学、図書館学を中心に発展したこの分野は、他の学問分野同様、1990年代のインターネットの登場により大きな転換を迎える。人文学でもデジタル化のプラットフォームが本格的に整備されるようになり、2006年には各国のデジタル・ヒューマニティーズの関連学会を束ねる連盟として Alliance of Digital Humanities Organization (ADHO) が発足。そして2000年以降には、それまで Computing and the Humanities や Humanities Computing など様々な呼称で呼

ばれていたものが〈デジタル・ヒューマニティーズ〉という名称に収斂していった。

近年、2012年以降では、インターネット上での画像データを公開・共有するための新しい規格であるIIIF (International Image Interoperability Framework〈トリプルアイエフ〉と略称される)を導入する研究図書館が急速に増えている。2015年にはオックスフォード大学で設立されたIIIFコンソーシアムには、ケンブリッジ大学、ハーバード大学、イエール大学、MIT、香港大学など53の機関がフルメンバーとして関わっており、日本からは東京大学、京都大学図書館機構、関西大学、国文学研究資料館などが参加している(2019年9月現在)。

IIIFは画像データの共有を円滑にするための規格である。これが策定される前は、個々のデジタル・アーカイブに蓄積された画像資料は相互運用性がなく、画像を閲覧するために別々に構築されたビューワーアプリケーションを使う必要があった。しかしIIIFに準拠することで、その煩わしさがなくなり、利用する側も公開のコストを大幅に削減できる。こうした流れから、今後はIIIFがデファクトの国際標準になると考えられている。

3 オープンアクセス・モデル

データや情報マネジメントとデジタル・スカラーシップ(ヒューマニティーズ)の歴史を概観すると、収集したデータを「他人が使う」ことを常に想定し、再利用・再配布がしやすい形式、それも国際的に共通したものを整えることに多大な労力が払われてきたことがわかる。それは、データを誰でも自由に使えて再利用でき、かつ再配布できるようにするという現代の「オープンアクセス・モデル」の基盤にもなっている。

Thomas Aquinasの著作のデジタル索引の制作に30年もの歳月を費やしたRoberto Busaも、研究に着手した動機を「他の研究者が使えるよ

うにするため」と述べている(Busa, Roberto, 1980: 83)。Thomas Aquinasの思想を言語学的観点から解明する研究によりグレゴリオ大学で博士号を取得したBusaは、当初、著作の索引を紙のカードで作っていたが、1千万ワード以上からなるThomas Aquinasの著作をカバーするにはなんらかの機械的手段が必要だった。それを可能にする技術を探して全米25以上の大学を訪ね歩いた彼がたどり着いたのが、電気機械式計算機の黎明期にあったIBMだった。当時IBMの社長だったThomas J. Watsonとの直談判により全面的な協力を得て、30年という歳月をかけてIndex Thomisticusは完成した。

1962年から67年の5年間には、60人以上の作業員をフルタイムで雇用し、今でいう「タグ付け」をしながら手作業でパンチカードに入力したという。パンチカードはその後磁気テープに取って代わられたが、25年間の入力期間で平均して1時間に2,200ワードも入力するという、気の遠くなるような作業が延々と続けられた。こうした作業は、BusaがComputing and Humanitiesにプロジェクトの回顧録を執筆した1980年の時点でも、コンピュータで6~7年はかかると推測されている(Busa, 1980)。光学文字認識(OCR)が普及した今では6~7日程度で終わってしまうかもしれない作業だが、この労力がすべて「他人が使えるように」という目的のために投じられたことに改めて驚かされる。

データが作成され、流通し、他人によって再利用されて新しい価値を生み出す「オープンアクセス・モデル」の具現化にインターネットが果たした役割の大きさはいうまでもない。インターネットは、データの流通と再利用の基盤インフラとして機能しているだけではなく、それ自身が、データを無数のユーザーが共有・再利用することにより発展してきた。公的資金を得た研究成果は学术界のみならず一般市民にも公開されるべきであるという規範も、インターネットの発達とともに確立していった。



このように、「オープンアクセス・モデル」は、研究成果やデータの発表形態、すなわち「アウトプット」側に大きな変化をもたらしたが、それと同時に「インプット」側、つまりデータの収集方法にも大きな変化をもたらしている²⁾。その一つとして、市民科学(別名クラウドサイエンス)が挙げられる。非職業研究者(一般市民)が観察者として研究の基礎データの収集に参加するとともに分析も行うのが市民科学(シティズンサイエンス)である。

これと同じような実践は、じつは古くから見られる。たとえば、鳥類学におけるパートウォッチャーによる標識調査、天文学におけるアマチュア研究家の貢献、生態学における環境モニタリングなどだ。現代の市民科学もこの伝統の延長線上にあるが、デジタル技術の発達によるアクセス性の向上やスマートフォンの普及によって、一般市民の参加形態は格段に多様化しつつある。

たとえば、「みんなで翻刻」プロジェクト³⁾では、古文書に記された歴史上の天変地異の記録を最新の科学に生かすために、インターネットを通じて一般市民約5,000人が翻刻作業に参加している。翻刻作業は従来、くずし字を解読する訓練を受けた専門家にしか行えなかったが、くずし字解読支援アプリを公開し、そこに搭載されたAIの手助けを受けながら、参加者はくずし字の解読方法を学びながら参加できる。

また、人間の脳や網膜神経をマッピングするEyewireというプロジェクト⁴⁾では、参加者が脳の神経経路を繋いでいくネットゲームが提供されている。神経の3D画像を見ながら神経接続している箇所を見つけていき、一つの経路をつなぐために要した時間や他のプレーヤーが描いた経路との一致度などをもとに、ポイントが付与されるというもの。他にも、数え切れない数の「市民参加型」研究プロジェクトが進行している。

ところで、市民という「非専門家」と職業研究者という「専門家」との区別は、特定の学範(ディシプリン)の制度化の前提なしにはありえない。

学範(ディシプリン)の制度化とは、「何を研究しているか」という研究対象による学問領域の境界があり、「いかにして研究するか」という理論的枠組や方法論が共有されていることが前提である。そして、大学などの高等教育機関における学部・学科および大学院で人材が育成され、学会などによって「誰が含まれるか」が定義され、さらには外部から「専門家」として承認され、権威づけられることを意味している。こうして制度化された学問は、専門家と非専門家とを区別し、多くの場合は前者が後者より上の立場にあるという上下関係があるとされる。そこには大きな倫理問題があり、その点については本特集の近藤論文に詳しい。

何のための学問か。社会に貢献する学問はどうあるべきか。そうした問いは、これからも真剣に議論されていくだろう。その議論において、デジタル・スカラシップの実践には大きな期待も寄せられている

4 情報を伝えるということ

もう一つ、デジタル・スカラシップがもたらす果実に、調査によって得られたデータの表現方法に新しい地平を拓いたことが挙げられる。社会調査の結果を多様なオーディエンスに〈伝える〉ということについては、本誌第19号の特集「調査と表現——伝えるための戦略」でも取り上げているが、社会調査の成果をわかりやすく表現し正確に伝えることは、調査の存在意義にかかわる重要なポイントである。

研究者の仕事は「活字にしてナンボ」とよく言われるが、デジタル・スカラシップには、そうした文字中心の「知」のかたちを再定義し、人間の全感覚を使ったものとして開放する可能性をもっている。五感のうちでも特に情報入力量が圧倒的に多いとされる「視覚」は重要な意味をもつ。デジタル技術を用いた視覚化を伴う研究実践も数多くあるが、本稿では、「記憶の解凍」

というプロジェクトを紹介したい。

「記憶の解凍」とは、渡邊英徳が取り組んできたプロジェクトの総称で、「デジタルアーカイブ／社会において“ストック”されている資料を“フロー”化し、コミュニケーションを創発することで情報の価値を高め、継承へのモチベーションを生み出していく営み」(渡邊,2018:1)である。ここで言う“フロー”とは、蓄積(ストック)された情報を「流す(フロー)」こと、つまり「伝える」ことを意味している。ストックされた情報はフロー化されて人々に伝えられることにより新しい価値を生む。「記憶の解凍」プロジェクトはデジタル空間に集められたデータの意味や価値を、地図(空間)にプロットすることでその内容を読み取りやすくすることや、過去の白黒写真を人工知能(AI)で色付けすることにより、できごとの「実相」⁵⁾をよりリアルに伝えることに取り組んでいる。

図1は、原爆投下直後の広島市街の写真や被爆者の証言などのデータをGoogle Earth上に重層表示した「ヒロシマ・アーカイブ」⁶⁾のトップページ画面である。カラー写真で掲載できないのが残念だが、中央に描かれた円は広島に投下された原子爆弾を表している。地球を見下ろす視点から日本へそしてヒロシマへとズームインし、この原爆が投下されるまでをアニメーションで表した後、広島市街の立体地図に被爆体験や当時の写真などがレイアウトされた画面が現れる。一つ一つの写真をクリックすると資料が拡大表示され、1945年当時の状況と、被爆者が、そこでどんな体験をしたかを知ることができる仕様になっている。

この「ヒロシマ・アーカイブ」はスマートフォンのアプリでも利用できる。アプリの「AR(拡張現実)モード」を立ち上げると、図2のように実際の街並みをカメラを通して眺めながら、資料を重層的に閲覧できる。「ポケモンGO」でポケモンがカメラを通して街中に出現するように、今自分がいる「ここ」という同じ場所で過去に

起こった事柄を、立体的かつ現実的に感じることがができる。

このように、ヒロシマ・アーカイブは散在していた被爆史料(原爆体験データ)を三次元空間にまとめ、過去と現在という時間軸を超えてデータ同士を関連付けることで、原爆投下というできごとをより多面的に捉えることができる。

ここで注意していただきたいのは、このプロジェクトの核となる原爆体験データは、いわゆるビッグデータではないということだ。ツイッターやWi-Fiログデータのように自動生成されたものではなく、地道な調査によって得られたものである。



図1 ヒロシマ・アーカイブのトップページ
http://hiroshima.archiving.jp/index_jp.html



図2 ヒロシマ・アーカイブ画面
http://hiroshima.mapping.jp/arapp_jp.html



ヒロシマ・アーカイブの前身であるナガサキ・アーカイブは、長崎新聞社の「私の被爆ノート」に収録された証言を元に作成された。長崎新聞社は資料をインターネットで公開することについての許諾を証言者から得ていたが、ヒロシマ版を作るために最初に注目した「被爆者証言ビデオ」(広島平和記念資料館収蔵)は、二次利用が許可されていなかったため、利用することができなかった。そんななか、広島女学院高校の教員と生徒が中心になり、被爆体験者のインタビュー収録を行い、渡邊が当時勤務していた首都大学の学生らのチームが作業を行うかたちでヒロシマ・アーカイブはスタートした。このように「多様な人々」が作る「運動体」を、渡邊は「記憶のコミュニティ」と呼び「記憶の解凍」の一つの柱と位置づけ、他者との対面的コミュニケーションを通して浮かび上がる事実、記憶、生成されるデータを掘り起こすことに注力している(渡邊,2013)。

渡邊は一連の仕事を「作品」と呼ぶ。新しい「知の営み」として注目したい。

5 特集論文の紹介

人文学から発信されるデジタル・スカラシップは、人類の知識の源泉としてのデータや情報の価値を引き継ぎながら、それを新しいかたちで組み直そうとしている。では、それはどのような組み直しなのだろうか。このような問題意識に基づき、本特集では4名に執筆をお願いした。それぞれの論文について簡単に紹介しておきたい。

特集論文2 リアルタイムでのデジタル・ヒューマニティーズと災害の記録——東日本大震災と津波のオーラルナラティブ・アーカイブ

(Flavia Fulco, David H. Slater, Robin O' Day)

宮城県、岩手県、福島県の計7地区における被災者のインタビューから抽出した数百のショートクリップをまとめたアーカイブスの成立過程とともに、方法論的課題や新たな応用への可能

性を紹介している。3.11の被災地域での学生ボランティアグループの支援活動が基盤となり、その一環としてはじまった被災者へのインタビューを記録する活動の展開とともに、そこから派生する方法論的、倫理的課題についても論じている。

特集論文3 社会科学と情報学の出会い——観光に関する社会調査とビッグデータ分析の併用の可能性(鈴木貴久)

千葉県香取市佐原地区で行われる佐原の大祭における来訪者への調査データと、会場付近に設置されたWi-Fiシステムのアクセスポイントにおける祭り期間中のログデータを併用することにより、来訪者数や居住地域のより正確な推定を行っている。また、社会調査によって得たデータと、いわゆるビッグデータとを突き合わせることで、それぞれの利点とともに分析する上での課題について考察している。

特集論文4 日本の地名・人名データベースから立ち上がる記憶のコミュニティー(山田太造)

人間関係ネットワークの解明にトピックモデルLDA(Latent Dirichlet Allocation)を適用した例を紹介している。ここでは16世紀戦国期の薩摩島津家の家臣である上井兼兼が記した、1574年から1586年までの日記を使用し、そこに登場する人名がどのようなコンテキストで出現し、そのネットワークが時間的、空間的にどのように変化したのかを考察している。歴史の人名・地名データベースを駆使して、デジタル化した日記の文章から人名および地名を抽出し、外交、合戦、行事などの「イベント」に誰と誰が一緒にいたのかという共起関係とその変化をたどることで、当時の人間関係を再構築する。

特集論文5 オープンサイエンスとオープンガバナンスの倫理的諸問題(近藤康久)

社会の課題を解決するための、研究方法の新しい概念を整理した上で、市民と職業研究者と

の関係に非対称性を要因とする倫理的課題を提起している。たとえば、職業研究者にとってのインセンティブである論文出版やアカデミアでの就職・昇進は、市民ボランティアにとってのインセンティブとはならないから、市民ボランティアにはきちんと対価を支払うべきだと論じている。市民が主体的に課題を解決するためには、市民と行政、企業、専門家は、理想的には対等な立場で課題解決に参画するシビックテックを提唱している。

6 おわりに

以上のように、本特集では、人文学から発信されたデジタル・スカラシップの成果を足がかりに、被災者支援のためのボランティア活動とそこから派生する研究活動、社会調査とビッグデータの交配、歴史的史料とデジタル技術、専門家と一般市民の関係など、いわば、知識生産システムの「境界」を超える試みについて紹介する。本稿では人間にとってのデータや情報もつ意味について、デジタル技術の登場でも変わらなかった点を、特に強調してきたかもしれない。データの収集と格納、そしてその共有のためのシステムは、基本的には研究者自身のためであるとともに、他人のため、つまり最終的には人類の「知」の前進のためにある。

冒頭でデジタル技術の発達は「知の革命」をもたらしていると述べた。データ処理技術の発達にともない、そしてインターネットの出現により、データや情報の形に構造的な変化も現れている。本稿で紹介した「視覚化」はその一つではあるが、より端的にいうと、私達が認識する「知識の総体」のかたちが〈ツリー構造〉から〈ネットワーク構造〉へと変化しつつあるのではないだろうか。

たとえば、図3と図4を見比べていただきたい。図3は、19世紀ドイツの生物学者エルンスト・ヘッケルが記した生物系統樹で、生物の類縁関

係を図式化したものである。図4は、大腸菌の遺伝子発現調節が相互に影響しあって調節されている様子を表した遺伝子制御ネットワークである。ヘッケルの系統樹は、当時知り得る限りの生き物の種類を分類し整理するにあたり、最も合理的な方法で、植物界 (Plantae)、原生生物界 (Protista)、動物界 (Animalia) と分けている。それらは別々に存在し、互いに影響することはないと考えられていた。しかし、現代の私達から見ると、いずれもDNAという設計図をもち、そして、ある特定の種 (図4では大腸菌) の中でも遺伝子は相互に関係をもっていると考えer方が自然なのだ。

人文学と調査を専門とする社会科学は、ヘッケルの系統樹の例でいうと、植物界と動物界ほども遠い存在のように思われていた。しかし、

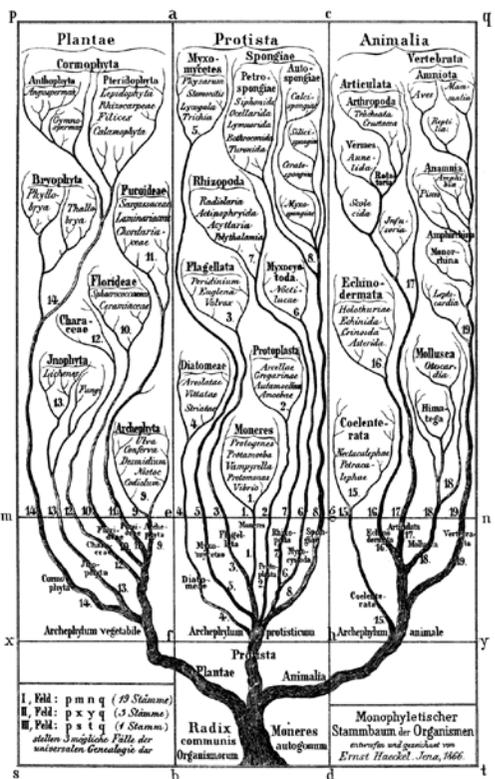


図3 ヘッケルの系統図
https://commons.wikimedia.org/wiki/File:Haeckel_arbol_bn.png

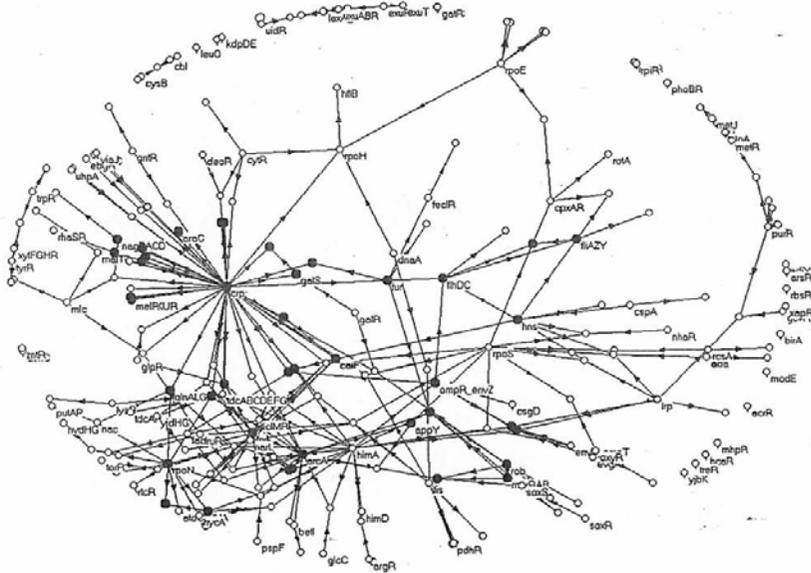


図4 大腸菌の遺伝子制御ネットワーク 出典:Alon 2006

生物をDNAレベルで解析するように、人文学と社会科学とをデジタル技術で接合すると、それぞれが美しい網の目のノードの一つとして立ち

現れてくる。

学問領域の境界を軽々と越境できる、そんな時代に私達は生きている。

注

- 1) 以下の記述は永崎 (2018) による。
- 2) この点について、本特集の近藤論文では「トップダウン」と「ボトムアップ」と表現している。
- 3) 京都大学古地震研究会によって2017年から始められたプロジェクト (<http://kozisin.info/>)。
- 4) <https://eyewire.org/explore>
- 5) 仏教の言葉で「一面的には捉えられない、真実の姿」で、原爆や災害の体験を伝えるという渡邊のデジタル・アーカイブを構築する原動力になった (渡邊, 2013)。
- 6) http://hiroshima.archiving.jp/index_jp.html

文献

Alon, Uri. ,2006, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman & Hall/CRC. (=倉田博之, 宮野悟 (訳), 2008, 『システム生物学入門—生物回路の設計原理』, 共立出版)

Burdick, Anne, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jefferey Schnapp, 2012, *Digital Humanities*, Cambridge: MIT Press. <https://mitpress.mit.edu/books/digitalhumanities>.

Blair, Ann., 2010, *Too Much to Know: Managing Scholarly Information before the Modern Age*, New Haven: Yale University Press.

Busa, Roberto., 1980, “The Annals of Humanities Computing: The Index Thomisticus” *Computers and the Humanities*, 14: 83–90.

- Pearce, Nick, Martin Weller, Eileen Scanlon, and Sam Kinsley, 2010, "Digital Scholarship Considered: How New Technologies Could Transform Academic Work" *Education*, 16 (1): 33-44, <https://ineducation.ca/ineducation/article/view/44>.
- ブリュノ・ラセル, 木村恵一・荒俣宏 (訳), 1998, 『本の歴史』創元社.
- 永崎研宣, 2013, 「人文学分野とサイバーインフラストラクチャー: デジタル・ヒューマニティーズにおける現状と課題」『情報の科学と技術』63 (9): 369-376.
- 総務省, 2018, 『平成30年版 情報通信白書』, (<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h30/html/nd102300.html>).
- 渡邊英徳, 2013, 『データを紡いで社会につなぐ: デジタルアーカイブのつくり方』講談社.
- , 2018, 「「記憶の解凍」: 資料の“フロー”化とコミュニケーションの創発による記憶の継承」『立命館平和研究』, 19: 1-12.



特集論文

2

リアルタイムでのデジタル・ ヒューマニティーズと災害の記録 東日本大震災と津波の オーラルナラティブ・アーカイブ

Flavia Fulco

東北大学
災害科学国際研究所 助教授

David H. Slater

上智大学
国際教養学部 教授

Robin O'Day

ノースジョージア大学
人類学部 助教授

1 はじめに

近年までデジタル・ヒューマニティーズの概念は、主に「図書館のデジタル化」と「文書に基づくアーカイブ資料」と解釈されてきた (Schreibmanほか,2016; Berry,2012)。しかし、現代史において重大な事象に関わる生存者による証言 (オーラルナラティブ) をはじめ、さまざまな“生きる資料”を用いたアーカイブ研究も含まれることで、この10年の間にデジタル・ヒューマニティーズの概念はさらなる広がりを見せている。具体的には、「デジタル社会科学」(Spiro,2014; Quinell,2012)、「オーラル・ヒストリー」(Boyd and Larson,2014)、「その他のアート形態のもの」(Travis and von Lünen,2016) に分類できる研究までもが含まれるようになり、デジタル・ヒューマニティーズは広く解釈されるようになった。類似した方法論で新たな応用に発展する可能性もあり、はるかに広範囲な現象も研究することが期待できる。これは、ある出来事が発生し、継続している最中でも、調査対象者がその出来事に関連する研究成果にアクセスできることも意味する。

研究サイクルの時間が短縮されることにより、研究者は従来よりも早く、事象と調査対象者に関わるようになった。また、直接に調査対象

者と関わることにより、研究者の政治的、民族的なポジショナリティ (立場性) も変化した。この変化により、調査対象者 (subjects) と研究内での表象 (representation) との関係性だけでなく、データ収集やアーカイブ作成にあたっての方針の再考が求められるようになった。デジタル・ヒューマニティーズは誕生して間もない分野だが、アイデンティティの課題とその機会、そしてベスト・プラクティス (模範的実践事例) への再考プロセスを始めるための前段階である。

本稿は、2011年3月11日に日本の東北地方の海岸を直撃した東北地方太平洋沖地震と津波 (以下、「3.11」) の被災者の口から語られた実体験を収集した、オーラルナラティブ・アーカイブ「東北からの声 (Voices from Tohoku)」の誕生と、その開発までの経緯を記したものである。日本において観測史上最大の地震が東北地方を襲い、それにより引き起こされた津波で被災地は大破し、2万人以上の犠牲者が出た (行方不明の人数を含む)。自然災害が頻発する時代において、研究者の役割はどうあるべきかを批判的に考え直す必要性、また、社会に (特に震災の影響を受けたコミュニティに) 利益をもたらす方法で行動をおこす必要性から「東北からの声」は生まれた。

自然災害と隣り合わせの時代に、研究者にはっきりと求められていること、社会に貢献できることはなんだろうか。私たちの研究・学問は、

現代社会とどのくらい関連があるのか。本稿で説明する方法論的・技術的なシフトは、上記の問いに対して部分的に答えるものとして開発された。本稿で特に注目すべき点は、「東北からの声」は学部生のためのサービラーニングプログラムのなかで実施されたことだ。本稿が、他大学が独自のデジタル社会調査のプログラムを開催するさいの参考資料となることを願う。

2 プロジェクトの始まり

日本では、1995年の阪神・淡路大震災の発生以来、災害救助のボランティアグループの活動が増加している (Hasegawa, 2007 ; Sakamoto, 2012)。2011年の「3.11」で、最も被害の大きかった3つの県 (福島、宮城、岩手) が、国内外の多くのボランティアから支援を受けた。そのなかには、上智大学の学生と研究者のグループもいた。オーラルナラティブ・アーカイブ「東北からの声」は、「3.11」の被災地域でボランティア活動を行った上智大学国際教養学部グループによってつくられたウェブサイトである。

当初、東北へ訪問した目的は、奪われてしまった被災者の日常を取り戻すために、手助けをすることだった。学生たちは、津波で流された倒木や泥をよけ、瓦礫を運び、また東北の沿岸地域で他の支援を行う合間に、被災地にある多くのコミュニティを訪れた。津波発生後のボランティアは特に困難だった。「東北からの声」の基盤となったボランティア活動のダイナミクスについては、Slater and Veselić (2014a) で説明している。オーラルナラティブ・アーカイブは、東北のコミュニティが求める支援を行うことを目的とした、ボランティア活動から生まれた。

ボランティアによる支援活動をもとにして、私たちは被災者を対象にインフォーマルなトークセッションを開催した。大学での講義や演習を通して、学生の民俗学の知識がつき、インタビューの技術的スキルが上達するにつれて、彼

らのインタビューはよりフォーマルなかたちへと発展していった。私たちは、語り手から研究同意書の署名を得た上で、インタビューをデジタル映像で記録した。それを知ると、多くの語り手は自らの被災体験を詳細に語った。

東北支援プロジェクトの多くは短期間の実施だった。いっぽう、本プロジェクトは社会的ダイナミクスを理解した上で、被災者をはじめとする重要な情報提供者との関係性を深めるべく、同じ地域を数年の間に繰り返し訪問した。各地域が直面している問題への理解と、被災者との関係性が深まるにつれて、インタビューの内容は複雑になり厚みが増した。これらの全てをデジタル映像に収めたことで内容を見返すことができる。これにより、撮影時には気づかなかった側面に気づいたり、物語の重要性を認識したりすることができ、その後のインタビューに生かすことができた。

プロジェクトが発展するにつれて、私たちは既存のインタビューを整理した。次のインタビューの方針を検討するために「グラウンデッド・セオリー (grounded theory)」という質的社会調査の技法 (Glaser and Strauss, 1967) を使い、「コード化 (coding)」して主題のラベルを作成した。年月を重ねるにつれてデータが蓄積され、デジタルアーカイブ・プラットフォームと様々な地域のサブカテゴリー間の互換性を備えた検索方法を採用し始めた。時には、録音された語りのコーパス全体にわたってテーマをさかのぼって検討できるようにするために、インタビューを再度コード化する必要があった。

このプロセスで作成したウェブサイト「東北からの声」には、7つのコミュニティ (宮城県石巻市雄勝町、名取市閑上地区、本吉郡南三陸町、石巻市、岩手県上閉伊郡大槌町、および福島県内の2地域) の語り手たちによるインタビュー動画から抜粋した、数百の映像クリップを掲載している。「東北からの声」は、500時間を超えるインタビューをストックした学術的アーカイブで、



一般に公開されているウェブサイトである。他の研究者も共有することが可能であり、累計8万件以上のアクセスがある。特筆すべきは、ウェブサイトへのアクセス元の大半が東北地方という点である。語り手の多くは、自分たちの映像クリップが公開されることだけでなく、同じように被災した他の地域の人々が、自然災害にどのように対処したのかに興味を示していた。

3 コミュニティ支援活動から生じる共同研究

すでに説明したように、本プロジェクトは震災直後の私たちのボランティア体験から生まれた。語り手の経験、悲しみに寄り添い、そしてニーズをより理解したことが、ボランティア活動を越えた手助けをする道を模索するきっかけとなった。社会科学の理論とその方法を実践する訓練を受けた学生たちは、ボランティア活動とは異なる、より価値のある方法としてスキルを利用できないかと考えた。しかし、危機が差し迫った状況で、習得した知識やスキルを応用することは、学部生には荷が重いかもしれない。正直なところ、本プロジェクトが始まった時、私たち教員はこうした問いへの答えがわからなかった。悩んだ末、とにかく前に進むために、学部生には本プロジェクトに積極的に参加してもらうことにした。

私たちが東北でボランティア活動をし、災害のオーラルナラティブを記録し始めた時、カメラを持つ私たちを見た被災者は「伝えたい話がある」という強い思いから、「撮影してほしい」と尋ねてきた。被災者が進んで取材を受けた理由はいくつもあるが、もっとも共通する気持ちは「メッセージを送りたい」というものだった。被災者がこの必要性に駆られたのは、他のジャーナリストや研究者からインタビューを受けたことも関連している。多くの人が、彼らのふるまいや取材した内容に不服や不満を感じていたからだ。

たとえば、一部のジャーナリストたちが、彼ら

が事前に用意したシナリオに合うように、被災者から現実の小さな断片を集めるようなかたちでインタビューをしていることに、被災者は不平を述べていた。レポーターは、話の「おいしいところ」に飛びついた後はその場を去り、それ以上のことはけっして聞いてこなかった。それを皮肉って、地域住民はジャーナリストを「カラス」と呼んでいた (Slater and Veselič, 2014b)。地域住民にインタビューを求める人たちは、住民たちの物語を多くの人に知ってもらえるように正当に扱うのではなく、ただ利用しただけなのだと感じていた。この批判は、状況の複雑さを見逃して調査しているように思われる研究者たちにも及んだ。「被災地で何が起きているのか」という概要を理解するには調査は欠かせないが、「事を単純化しすぎなのだ」と地域住民たちは口をそろえた。調査にきた研究者たちからしばしば、被災者の感情を数値で表現するようにと求められたことを嘆いていた。「被災体験の物語をすべて聞かせてほしい」と言われることは、ほぼなかったという。

私たちが何よりもまず被災者の手助けをする目的でやってきて、メディアやほかの研究者が聞きたがらない被災体験に興味を示していることがわかると、被災者の人びとは私たちを信頼し、彼ら自身の言葉で被災時のことを語ってくれた。

4 新しいインタビュー技法の開発

私たちのインタビューの方法は、取材を受けた被災者がこれまで経験したものとは異なるものだった。ローカルな文脈に注意を払うと、どのコミュニティでも似たような問題を抱えているが、それぞれの文脈は極めて特殊であることが理解できた。

「3.11」の災害を一連の悲惨な出来事ととらえられがちだったが、各地域の状況を取材すると震災のリアリティはそれぞれで異なり、複雑であることがわかってきた。津波は、多くの地域を

含む、およそ400kmに及ぶ沿岸地域（青森県から茨城県まで）を襲った。さらに、福島第一原子力発電所における原子力発電所事故は、津波による被害とは別に、対処すべき多くの特殊性をもつ二次被害を引き起こした。私たちが地域の人々に尋ねた質問のひとつは、「あなたがここで何を経験しているか、私たちに説明していただけますか」だった。他の研究者たちが行っていたインタビューと、私たちが採用する「民俗学的インタビュー」の最も大きな違いは、特定の答えを探すよりも、地域住民の心理的状況をより正確に理解できる問いを探そうとした点だろう。

私たちのインタビュー（通常1人につき60分）は半構造化面接で、生存者の個人的な経験から始まり、地域の過去・現在・未来にわたる、次の3つの広範な質問のみに絞った。

- 震災以前のあなたのコミュニティについてお話しいただけますか。
- 震災から今日までのあなたの体験についてお話しいただけますか。
- 未来に期待するものについてお話しいただけますか。

このような自由回答の質問形式をとることで、語り手が自身の方法で物語を語り、相手に話を聞いてもらっていると感じられるようになった。インタビュアーとして、私たちはそれぞれの物語を注意深く聴き、物語が展開した時には、明示的な質問をするようにした。被災者の状況・心情を正確につかめる〈よい質問〉を探して調査フィールドに踏み込むことで、エスノグラファー（民族誌学者）は地域住民にとって大切な要素がなにかを認識できた。いっぽう、語り手たちは自分たちの状況を正確に伝える方法を研究者に教えることにより、自分たちが主体性をもっていると感じることができた。語り手とインタビュアー／研究者の間の潜在的な信頼関係を築くには、ある程度の時間が必要になる。つまり、

私たちのこの方法論はデータ収集のために、長期間にわたって地域と関わることを要する。

また、この研究プロセスは、地域的文脈に足を踏み入れる前に、「何がコミュニティにとって最善か」を〈専門家〉は知るべきだという立場をとった。これまでの研究プロセスの常識を打ち破るものである。

5 デジタル記録から コミュニティ・アーカイブへ デジタル社会科学

私たちが地域社会に深く関わると、語り手たちの多くから「この資料を使ってあなたたちは何をするつもりなのですか」と、プロジェクトの根幹に関わる質問を投げかけられるようになった。語り手たちはこのプロジェクトの短期的な目的を理解していたが、プロジェクトの長期目標とねらいも知りたがっていた。「インタビューはどのように、この状況を改善してくれるのですか」と聞かれたが、私たちは彼らが納得できる答えを出せなかった。おそらく今でも〈正しい〉答えを出すのは難しいだろう。

はじめから私たちが理解していたことは、多くの人々が悲劇に対処しようとする物語をたくさんもっていることだった。被災者の多くが家や仕事を失い、さらに不幸な人は家族も失った。もし、こうした物語を収集しなければ、人の目に触れられないままであることはわかっていた。研究者としてこうした体験を収集し、保存することの必要性を理解しながら、当時は地域住民へ私たちができる感情的な支援と実用的な手助けに焦点を当てていた。

当時の出来事にタイムリーに対応し、研究者と地域コミュニティとの共同作業の成果を示したにもかかわらず、アーカイブの存在はプロジェクトの長期的なねらいと必要性を今なお問い続けている。「東北からの声」を始めた時に語り手の多くが指摘したように、「3.11」に巻き込まれたコミュニティに何が起こったのかを文書化することは、アーカイブ制作にあたっての主な動機



のひとつだった。研究を行ったコミュニティ全体で、「何が起きたのか」を覚えておくために記録することは重要であるという認識は共有されていた。

『Oral History and Digital Humanities』の編集者は次のように指摘している。「記憶で形成され、ナラティブによって語られる物語は、共鳴や関わり合いをもたらすか、記憶に刻まれて留まるか、あるいは時間とともに忘れ去られるかもしれない。歴史は、時のなかで保存された断片に基づいた人類の物語によって成り立っている。物質文化（建築、美術、壊れた土器）、文字で書かれているもの（日記、記録、本）は、歴史を理解したいと思う者のために形ある遺産を残す。その一方で、物語として表れる話し言葉は、従来、保存することがより難しいとされてきた。しかし、歴史家が求めているものは声であり、目撃談であり、そして学者がじかに質問し、過去の語られた答えに取り組むことのできる特権と機会である」（Boyd and Larson, 2014）。

実際に、「3.11」について書かれ、オーラルナラティブを主として集めた優れた書籍は多数ある（岩上, 2012, 2014; 金菱, 2012）。そうした書籍はインタビュー記録の文字起こしと、編集をとおしてできたものであったり、当事者が自分の手で体験をつづったものだったりした。しかし、本プロジェクトのデータはそうしたものは異なる。この違いこそがデジタル社会科学が秘めているポテンシャルのひとつである。映像は、記述された文字よりも説得力を持つ媒体だ。私たちはすべての物語を映像でアーカイブ化する機会を得たことで、新しいデジタル技術によって生み出される可能性を見出すことができた。

しかし、こうした物語を公開することで何ができるのだろうか。災害の物語を保存・記録することは大切だが、それだけで十分なのか。その情報によって、もっと何かできることはないのか。最初からデジタル映像でインタビューを記録したことで、新たなチャンスを得たが、同時に

課題も残る。そのひとつは、物語にとって〈意味のある公開方法〉を見つけることが挙げられる。地域との協働プロセスは容易に達成できるものではなかった。地域の固有性による違いや分断、痛みを伴う記憶など、語り手の複雑な感情や反応に配慮する必要があった。最初にインタビューした多くの人びとにとって、被災したときの記憶を思い起こすことは難しいものだった。時が経つにつれ、地域住民の声を集めたアーカイブは役立つようになった。アーカイブによって、地域住民は何が起きたかを異なる視点で思い出し、その意味を付与するようになったのだ。プロジェクトが発展し続けるうちに、私たちはこのプロジェクトをより多くの人たちにも届けることと、地域住民の参加の両方を望むようになった。

6 語り手のエンパワーメントと公共発信

時間が経過するにつれ、このプロジェクトを被災者と共同作業することにより、私たちはやりがいをを感じるようになった。いっぽう、コミュニティの人びとも研究に対し能動的な役割を与えられることで、主体性を感じる事ができた。前述したように、地域コミュニティと協力し合うことは、公的アーカイブをつくるうえで不可欠だった。

この過程で、500時間を超えるオーラルナラティブのインタビューが収集された。当事者には、それぞれのインタビューを文字起こした原稿のコピーを渡した。語り手たちは私たちに、「災害が他の地域にもどのような被害をもたらしたのかをもっと知るために、それぞれの地域を調べられるようなものをつくってほしい」と求めた。その言葉によって、ウェブサイトからアクセスできる形で公的アーカイブをつくることは、語り手のニーズを満たすことにつながる事が明白になった。

最初のうちは、こうした資料を素材として、ウェブサイトをどの程度正確につくり出せるか

を把握できなかった。500時間にも及ぶオーラルナラティブのインタビューを、すべて公的アーカイブにアップロードすることは、動画の数は圧倒的だが、もしかすると本来のねらいと逆効果になる可能性があった。その資料の中には、公開するにはあまりにもデリケートなものもあったからだ。さらに、アーカイブのデータサイズが大きいため、ユーザーが利用しやすいウェブサイトを編成する必要があった。そこで私たちは語り手と相談しながら、長いインタビューから1~2分程度の長さの短い映像クリップに編集して、アップロードすることを決めた。それぞれの映像クリップは、語り手が伝えたい明確なメッセージが込められていた。

さらに、それぞれの映像クリップを“コード化”または“ラベル付け”も行った。たとえば、「放射線」、「再建」、「メディア」、「政府からの情報」などのラベルがある。パブリック・エスノグラフィーの提唱者たちは、「人類学の研究は理解しやすい方法で様々な人々に効果的に伝えられるよう変革されなければならない」ということを強調している。ほとんどの人類学の研究は専門的すぎるか、難解か、それゆえに資料を公共利用することを難しくし、一般人にとっては非生産的で不可解な散文で書かれている。効果的な伝達はパブリック・エスノグラフィーの重要な要素である。大学での講義と学会発表に加え、(本稿のように)エスノグラフィー研究を学術誌と書籍とで発表することにより、専門的で学術的な訓練は研究者をアカデミーの特権へとむかわせる。

本プロジェクトは、アカデミーという狭い範囲を超えて、パブリック・エスノグラフィーのひとつのあり方を提示したと自信をもっていえる。当初から、このデジタルアーカイブプロジェクトは、本調査を一般公開してインパクトのあるものにする試みであった。より広く社会に情報を発信するため、異なる伝達手段と新技術を継続的に試していくつもりである。公的デジタルアーカイブとしての「東北からの声」は、学界では珍し

い、だれでも利用可能なオープンソースである。

近年、研究のための資金獲得はますます困難になってきており、研究者が自らのデータを管理し続けることなく、他の研究者に“譲り渡す”ことが容易でないことは研究者として理解できる。しかし、さまざまな研究者が本プロジェクトの語り手たちの「声」を用いて、このプロジェクトを豊かにし、本研究で出した成果以上のものをコミュニティにもたらすことを信じている。プロジェクトの協働的な側面こそが、本研究の特徴の本質である。そのため、本研究では共同研究を望む研究者をやや独特な方法で招きたい。アーカイブに調査を追加し寄稿することに同意した研究者には、アーカイブ全体へのアクセス権を付与することになっている。この取り決めの背景には、アーカイブの成長を促進させ、学術的・公的な研究成果の増加につなげたいという考えがある。

「東北からの声」プロジェクトについて言及する際、2つの異なるアーカイブを参照していることをここで注記したい。ひとつは、誰でも自身のコンピュータからアクセスできるウェブサイトである。公的アーカイブは、語り手の協力を得て編集したため、だれでもアクセスできるものとしている。もうひとつは、未編集のインタビューを含む非公開のアーカイブである。上記の公的アーカイブに加え、新たに作成された。共同研究の利点を理解し、アクセスを求める研究者ならば、この規模の大きいアーカイブを利用することができる (David Slater: dhslater@gmail.com にご連絡いただきたい)。

7 方法論のおよび学術的課題

前節でデジタルアーカイブプロジェクト「東北からの声」がどのようにして生まれたか、そしていまもなお存在している重要な側面について述べた。今から数年後(あるいは数十年後)、最終的にオーラルナラティブ・アーカイブはどう



役立つのだろうか。さらに、将来に起こり得る、災害の悪影響を緩和するのに役立てることができのだろうか。また、大惨事を生き延びた人々にどう役立つのだろうか。

オーラルナラティブとはどのようなデータなのか。いわゆる「ビッグデータ」のアーカイブに強い関心をむけている社会の潮流に、オーラルナラティブのアーカイブも乗っているように思える。行動分析やアルゴリズムを介して分析される「ビッグデータ」とは異なり、オーラルナラティブは均質化した分析に対抗する。しかし、物語は、コミュニティの集合的記憶に影響を与えるだけでなく、再建されたコミュニティのアイデンティティを強化する勇敢さ、団結、そして救済を強調する「集団主義的」物語のなかで、その持ち味が均質化される危険性をもはらんでいる。このプロセスを通して、個々の経験は平坦化、あるいは沈黙化させられ、最終的には忘れ去られてしまう。これは部分的に自然な過程であり、コントロールするのは非常に難しい。原点となる物語に立ち戻り、時間の経過とともに無視され得る問題のいくつかに耳を傾ける機会があることは有用である。

アレックスandro・ポルテッリ (Alessandro Portelli) によれば、オーラルナラティブを特別なものにしてしているのは、それらが口述である点だという (Portelli, 1998: 64)。ポルテッリがオーラル・ヒストリーの特徴について初めて記した際、オーラルデータの収集結果を要約・伝達する方法として通常用いられるトランスクリプションと比較して、声を記録することの特異性を考察していた。「声」の特異性は、最近では映像を含めればさらに重要な要素となっている。

1980年代のビデオ・オーラル・ヒストリーの先駆者たちは、映像技術を使うことの唯一の問題が映像技術のコスト (Wilson, 1986: 33) と映像テープの劣化を防ぐ難しさ (Gardener, 1984: 107) だと懸念していた。しかし今日では、技術の進歩によってこうした問題は克服され、インタビュー

を映像化することのメリットや利点に注目することができる。データを収集する瞬間、語り手のなかにはカメラをむけることによって「一時的に自信が増す」人たちもいる (Wilson, 1986: 33)。潜在的に広いオーディエンスにむけて話していることから、自分たちの話は重要で語る価値があると思えることで、一部の語り手は主体性 (empowerment) を感じられるからかもしれない。

映像化したインタビューでは、声のトーンや会話のリズムだけでなく、映像以外の方法では永遠に失われるであろう顔の表情やボディランゲージも考察できる。プロジェクトのデータ収集段階では、インタビューの多くが語り手の家で記録された。記録時は、語り手が主体的にリラックスして話し出せるように、彼らがどうしても伝えたいと思っていることを質問した。話が活発に展開されるように、インタビューの多くは一对一ではなく、同時に複数の語り手 (通常は家族) とともに行われた。こうした環境を整えて、ビデオカメラの使用することで、語り手にとって重要な要素をつかめるだけでなく、また、インタビューが行われている場所の重要性も捉えることができる (Charlton, 1984)。

一例を挙げれば、仮設住宅で暮らすある夫婦のインタビューをしていた際に、背後でパチパチという音がしていた。その音はガスコンロの上にある、やかんから発せられたものだった。「もし誰かが訪ねてきたら体を暖められるように、熱い飲み物を出してあげたい。だから、お湯を沸かし続けておく必要があるの」と夫婦は説明した。そのため、録画された映像にその音が残ることになっても、夫婦はやかんの火を止めなかった。このようなふるまいは、仮設住宅での生活を余儀なくされている人々にとって、仮設住宅での狭苦しい生活のなかでいかに社会関係に配慮するかに注意が払われているかを示している。この映像のおかげで、インタビューの内容だけでなく、当時の仮設住宅の状況についての

洞察も得ることができる。

8 デジタル社会科学への新たな応用

インタビューが現場でスムーズに収集できるとは限らず、先に述べたように、人びとが発言をする準備ができていても限らない。「3.11」が引き起こした危機的状況では、非常に多くの地域社会が混乱し、メディア及びソーシャルメディアによる報道が広範囲に及ぶなかで、多くの被災者にとって直接的に自分たちの言葉を伝えられることが大切だった。被災地で何が起きているのかを自分自身で見聞きしたことで、その状況がようやく納得できたという声を何度も耳にした。また、災害の特質（何らかの形で、東北に住む誰もが被害を受けた国家的災害）のために、語り手は喜んで取材を受けたいとさえ思っていた。語り手たちは、自分たちの声が記録され、人々に届き、共有されることを望んでいたのだ。

状況は他の文脈のなかで変化し、この3年の間で「生の声」プロジェクトは東北から東京へと移った。このプロジェクトの第2段階では、初期段階からの変化はいくつかあったものの、未解決の新たな課題があった。

本研究の焦点は反原発運動に参加した若者にむけられた (Slaterほか,2015)。その若者たちは自分たちの大義をととても強く感じ、大衆デモというかたちで自分たちの意見を発信した。しかし、「活動家」として自分たちが社会のなかで「破壊的で異質でさえある社会構成員」とみなされるであろうことを認識していた。ほぼ全ての主要メンバーに、東北の被災者と同じ方法でインタビューを行ったものの、多くのメンバーは個人的な感情やそれぞれの物語を公的アーカイブに載せることに躊躇した。

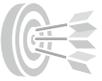
また、福島 (Slaterほか,2014c) からの多くの自主避難者、特に子どもを持つ若い母親にインタビューを行った。のちに反原発運動が復活してからは、この声を入れるためにアーカイブが

拡大された。多くの母親は、子どもを守るために夫や家族から離れた恐れからくる、二重のステイグマ (風評被害) を乗り越えようと苦闘していた。実際に彼女たちの多くが福島を離れたことを非難されてきた。また、新しい目的地に着いてからは、「汚染地域から来た」というステイグマを恐れており、原発事故の影響を受けた地域から来た人たちへの風評被害は、今なお強いままである。

こうした文脈のなかで、「インタビュー疲れ」に苦しみ、自分たちの話を誰にも語りたがらなかった東北地方の住民たちの一部とは異なり、自主避難した母親たちは「まだ誰にも話せていない」ために、インタビューを受けることを強く望んだ。それでも、自分たちの身元はもちろん、何らかの形で身元が明かされてしまう可能性があるような情報さえ、プロジェクトで共有したがないことがしばしばあった。活動家や母親たちのプライバシー保護の方法が当事者から承認され、その体験を信頼できる情報として伝えられる方法を見つけれない限り、活動家や母親のアーカイブはまだ公開されない。

プロジェクトの第3段階に進むと、調査対象者は路上生活者や最近では日本の難民・難民申請者などにも及んだ。東京近郊の現代都市のリアリティに目をむけ、震災後の状況からはさらに踏み込んだかたちとなる。

ここ最近では、特にプライバシーの問題に関して、新たな課題に直面している。プロジェクトのデータ収集の時点で語り手の身元を保護するのは簡単だが、路上生活者の声は「東北からの声」のようにアーカイブを公開ウェブサイトで開示する方法を見つけることが難しい。難民申請者である語り手の多くは、本研究で自身の話を共有することに積極的である。16ヶ月の調査で、すでに50人以上の語り手に複数回のインタビューを行い、150時間以上に及ぶインタビューのデジタル記録を集めた。このプロジェクトは「日本社会への関心を高める価値がある」と多くの



語り手は考えているが、同時にその結果として公衆の目に身をさらすことを恐れている。しかし、私たちは難民申請者の語り手のおよそ半分から、映像使用の許可を得ることができた。

収集やアーカイブ、その配布方法の詳細などは、プロジェクトやそのトピック、規模によって異なるだろうが、本稿では「デジタル・スカラーシップ」と呼ぶものの可能性と課題について概説してきた。方法論的・技術的な課題だけでな

く、取り組むべき学術的・倫理的な問題も含まれている。それでも、こうした課題は解決する努力に値するものである。頻繁に引用される学術論文よりもはるかに広範な読者へ届けられる可能性、あるいは、マスメディアが制作したほとんどの作品よりも誠実で直接的な方法で当事者の経験を捉える可能性があるからだ。本研究は、今後もこのアプローチを発展させていくつもりである。

文献

- Berry, D.M. ed., 2012, *Understanding Digital Humanities*, London: Palgrave Macmillan.
- Boyd, A.D., M. A. Larson eds., 2014, *Oral History and Digital Humanities*, New York: Palgrave MacMillan.
- Charlton, T. L., 1984, "Videotaped Oral Histories: Problems and Prospects", *The American Archivist*, 47(3): 228-236.
- Gardner, J., 1984, "Oral History and Video in Theory and Practice", *The Oral History Review*, 12: 105-111.
- Glaser, B.G., A.L. Strauss, 1967, *The Discovery of Grounded Theory: Strategies for Qualitative Research*, New Brunswick & London: Aldine Transaction. [reprinted in 2006]
- Hasegawa, K., 2007, "Volunteerism and the State in Japan", *The Asia-Pacific Journal: Japan Focus*, 5(12): 1-5.
- 岩上安身, 2012, 『百人百話 第一集』三一書房.
- , 2014, 『百人百話 第二集』三一書房.
- 金菱 清, 2012, 『3.11 慟哭の記録——71人が体感した大津波・原発・巨大地震』新曜社.
- Portelli, A., 1998, "What makes Oral History different?", R. Perks & A. Thomson, *The Oral History Reader*, London & New York: Routledge, 63-74.
- Quinnell S.L., 2012, "Digital Social Science Vs Digital Humanities: Who Does What & Does It Matter?", *Social Science Space*, (Retrieved April 06, 2019, <https://www.socialsciencespace.com/2012/07/digital-social-science-vs-digital-humanities-who-does-what-does-it-matter/>).
- Sakamoto, M., 2012, "The Rise of NGOs/NPOs in Emergency Relief in the Great East Japan Earthquake", *Japan Social Innovation Journal*, 2: 26-35.
- Schreibman, S., R. Siemens & J. Unsworth (eds.), 2016, *A New Companion to Digital Humanities*, UK: John Wiley and Sons.
- Slater, D.H. & M. Veselić, 2014a, "Public Anthropology of Disaster and Recover 'Archive of Hope' (希望アーカイブ)", *Japanese Review of Cultural Anthropology*, 15: 115-126.
- , 2014b, "Voices from Tohoku: 'Public' Research, New Media Practices and the 'Archive of Hope'", 5: *Designing Media Ecology*, 1: 28-41.
- Slater, David H., Rika Morioka & Haruka Danzuka 2014c, "Micro-Politics of Radiation", *Critical Asian Studies*, 46:3, 485-508, (DOI: 10.1080/14672715.2014.935138).
- Slater, D.H, R. O' Day, Uno S., L. Kindstrand, Takano C. 2015, "SEALDs (Students Emergency Action for Liberal Democracy): Research Note on Contemporary Youth Politics in Japan", *The Asia-Pacific Journal: Japan Focus* 13(37) 1-26.
- Spiro L., 2014, "Defining Digital Social Science" dh+lib, (Retrieved April 6, 2019, <https://acrl.ala.org/dh/2014/04/09/defining-digital-social-sciences/>).
- Travis C. & A. von Lünen (eds.), 2016, *The Digital Arts and Humanities: Neogeography, Social Media and Big Data Integrations and Applications*, Switzerland: Springer.
- Wilson, J., 1986, "The Afro-American Labor Leadership Oral/Video History Series", *The Oral History Review*, 14: pp. 27-33.

3

社会科学と情報学の出会い 観光に関する社会調査と ビッグデータ分析の併用の可能性

鈴木貴久

津田塾大学総合政策学部 特任助教

1 はじめに

本稿では、社会調査とビッグデータの組み合わせでどのようなことが可能になったのか、近年注目が集まっている観光統計を基に概説する。前半では、既存の社会調査のみを用いた観光統計の収集に限界があることと、各種ビッグデータを併用することの有効性について論じる。後半では、関東三大祭りの一つ、千葉県香取市の「佐原の大祭」への来訪者数やその特性などについて、社会調査とWi-Fiシステムログデータを併用して分析した事例を紹介しながら、その併用の可能性と課題について論じる。

従来、観光に関する調査や研究は、文化的側面、地域経済や来訪者の心理的効果などを対象として社会科学の領域として扱われ、そこでは社会調査によるアプローチが中心であった。しかし、近年では観光統計や観光情報といった部分において、ビッグデータを用いた情報学的な取り組みが急増している。2020年東京オリンピック・パラリンピック競技大会の開催やそれに伴う海外からの来訪者の増加を踏まえ、観光は国の成長戦略の柱、地方創生への切り札として見なされている。その中で制定された観光立国推進基本計画¹⁾では、地域観光における実態を把握し、施策立案へ利活用することの重要性が示されている。

なかでも特に重要となる観光統計は、来訪者数や行先、観光地における消費額等の情報を収集するために実施される統計調査であり、来訪者の「規模」や「性質」を的確に把握することが目的である(宮川, 2009)。ここでの規模は来訪者の人数や消費額などの大きさに関する情報であり、性質は来訪者の特性や消費の種類に関する情報を含んでいる。

これらの観光統計データの取得に際し、従来は社会調査を主な手法としてきた。たとえば、国土交通省は以下のような調査を行い、各種観光統計を収集している。

- 「宿泊旅行統計調査」²⁾では、宿泊施設を対象にして宿泊客の数を調べている。
- 「旅行・観光消費動向調査」³⁾では、住民基本台帳からの無作為抽出による郵送調査によって、旅行に行った回数やそこでの消費額などを調べている。
- 「全国観光入込客統計」⁴⁾では、都道府県ごとの来訪者数や特定の観光スポットやイベントへの来訪者数をとりまとめている。

これらの公的統計はいずれも、来訪者数や消費額など、観光の規模についての測定が主な目的である。

一方で、来訪者の性質に関する統計の収集についても社会調査が主な手法であった。たとえ



ば、観光地への愛着度やリピーターの存在は、持続的な活性化のために不可欠であり、その実態や関連要因を把握することは極めて重要となる。社会調査を用いた一連の研究では、満足度や愛着度などリピーターとの関連 (Devesa et al., 2010; McDowall, 2010; 岡野ほか, 2018; Xu & Fox, 2014; 弓田・原田, 2015) や、消費行動との関連 (Jurdana & Frleta, 2017; 鈴木・藤井, 2008) について検討している。

2 社会調査のみに頼ることの限界

しかし、社会調査のみを用いた観光統計の収集には限界がある。たとえば、都道府県単位の旅行者数については、上記複数の公的統計を用いた推定が行われているが、調査ごとに対象や手法が異なることによる数値の違いが生じており、その精度には改善の余地がある⁵⁾。また、「全国観光入込客統計」では、各都道府県内の主要な観光地やイベントの管理者から、来訪者数を報告してもらっているが、その測定方法についても大きな課題がある。というのも、入場券などが必要な施設やイベントの来訪者数は正確に把握できるが、街中で行われる通過型イベントなどでは、来訪者数を把握するための推定が行われているからである。

その簡易的な方法として、「一定範囲を定めて、その範囲内を最盛時間に移動しながら利用者を数え、これに回転数と全面積をかけ、さらに一定範囲の面積で割る」⁶⁾という手順が提示されている。しかし、最も人が集中した時間帯や区域の人数を等倍するという方法では、時間帯や場所によって人数が異なることなどが考慮されておらず、過大に推定され、実数からの乖離が生じる⁷⁾。

観光地ごとの来訪者数は観光統計として最も基礎になるデータであるが、社会調査をベースとして得られた値は実数から乖離している可能性が高く、施策の根拠として用いるには限界がある。そのため、社会調査に頼るのだけではなく、

他に補完や代替となる手法が必要になる。

3 ビッグデータ分析の活用

社会調査の補完・代替的手法として、ビッグデータと呼ばれるデジタルデータの利用が挙げられる。調査対象者の負担やコストを抑えられるうえに多様なデータが得られることや、実時間のデータ収集とオンラインフィードバック制御ができるなど、ビッグデータを用いる利点は多岐にわたる。

社会調査では対象者に回答を積極的に依頼して、対象者がそれに応じることでデータが収集される。そこでは、回答率の低下とそれに伴うバイアス⁸⁾や、回答内容の歪みなどの課題が生じている。特に、自己回答形式の社会調査では、自分が調査の対象になっていることを意識することで回答が社会的に望ましい方向に歪むこと (Phillips & Clancy, 1972) や、過去の行動などを想起する際に記憶のズレが生じることも知られている (Boase & Ling, 2013)。

一方、ビッグデータは対象者自身が所有しているスマートフォンなどの端末を通じて収集される。協力する意図がある調査対象者が、普段どおりに行動するだけでデータ収集できるため、調査対象者の負担は低く、得られるデータの歪みなども低減される。さらに、標本抽出による社会調査では、母集団の規模の推定はあまり想定されていないが、ビッグデータ分析では母集団の規模の計測や推定も行われている。そのため、これまでは多大なコストと時間をかけて集めてきた公的統計データでさえも、短時間での収集が可能となっている (Blumenstock et al., 2015; Gebru et al., 2017)。

こうした利点をふまえて、観光統計の収集においてもビッグデータを用いた取り組みが進み、来訪者数の測定でも中心的手法となりつつある (e.g. Chen et al., 2016)。特に、観光地やイベントの来訪者数を測定するには、ある地点に存在

する人数の計測や人の移動の追跡などが必要であり、以下のようなデータや分析手法が用いられている。

携帯基地局データ

携帯電話が通信している基地局の情報から、携帯電話の位置情報が得られる。この情報を利用することで、あるエリアに滞在している人の数が計測可能になるが (Gonzalez et al., 2008), 位置情報の精度に関しては基地局の密度に依存するという欠点も存在する。都市部のように基地局が多いエリアでは数百メートル程度の誤差に抑えられるが、基地局が少ない農村部ではキロメートル単位で誤差が生じることもある。そのため、比較的広範囲での人の分布の把握には適しているが、農村部の観光地レベルでの人数の把握などには適していない。

GPS データ

GPS (位置情報) データを用いた観光に関する行動調査なども行われている。GPSは誤差10メートル以下の精度で端末の位置を追跡できるので、所有者の移動ルートや各地点での滞在時間なども計測可能である。

近年では多くの人がGPS機能を備えたスマートフォンなどの端末を所持しているが、そのデータを収集するには専用のアプリのインストールなどが必要になる。そのため、GPSデータを用いて分析を行う場合、調査対象者を募集してアプリをインストールしてもらったり、記録用端末を貸し出してデータを収集することが多い。このように、社会調査と同様のプロセスは必要になるが、調査期間中の移動は自動的に記録されるので、従来の社会調査よりも調査対象者の負担は少なく済む。GPSデータを用いることで、観光地における歩行者の行動を可視化し、滞留する地点などを明らかにできる (野村・岸本, 2006)。また、山本ら (2006) はGPSデータと社会調査を併用することで、新宿御苑に初めて来

訪した人とリピーターの間で行動パターンに違いが見られることを示している。さらに、Bohte & Maat (2009) は、一週間分のGPSデータを収集し、各移動における移動速度などから移動手段を推定し、歩行と車と自転車については7割程度の精度で推定できることを示した。そこから日常生活における移動手段ごとの利用頻度を整理して公的統計と比較したところ、近い値が得られることを確認している。

ソーシャルメディアデータ

SNSなどのソーシャルメディアも有益なデータとなる。投稿内容のテキストを分析することで、観光地など特定の場所にいるかどうかや、イベントに参加しているかどうかも推定可能である。de Liraら (2019) は、音楽イベントの開催前後を含めたツイッター投稿内容を分析し、各投稿者がイベントの参加者かどうかを推定している。その推定結果と人手で確認した結果を比較することで、7~8割程度の精度で推定できていることを示している。また、ソーシャルメディアでは投稿の際に位置情報 (緯度と経度) を追加する機能が存在する。Abbasiら (2015) は、シドニー周辺の位置情報が付された投稿を集めて、住民と観光目的の来訪者の識別を行った。ソーシャルメディアへの投稿が自宅や職場から多くなされることを考慮し、同一の位置から複数回投稿している人は「シドニー在住者」、それ以外を「来訪者」として分類した。住民と来訪者の投稿の地点を比較することで、住民は住宅街からの投稿が多く、来訪者は観光地などからの投稿が多いことも示している。さらに、緯度と経度以外の位置情報として、投稿場所の名称を投稿時に付与することもできる。CaféやStoreなどの単語を含むような位置情報データと投稿内容を組み合わせて分析することで、投稿者の行動パターンをより詳細に分類することも可能になる (Hasan & Ukkusuri, 2014)。



Wi-Fiシステムログデータ

Wi-Fiアクセスポイントのシステム履歴（ログ）データを活用して、人々の動きを計測する手法も提案されている（Sonehara et al., 2019）。この手法では、スマートフォンなどの端末がWi-Fiアクセスポイントと接続する際に交換されるシステムログを用いる。この情報は、Wi-Fiを通じてインターネットにアクセスしなくても、アクセスポイントから一定の距離内に入ると記録される。アクセスポイントの場所と端末とを関連付けることで、端末の位置情報を計測している。また、端末の個別識別が可能であり、同一人物が複数回通過した場合などでも実人数を計測できる。さらに、Wi-Fiサービス提供事業者は、契約者の年齢、性別、居住地域などの登録された属性情報と結びつけることも可能になる。

曾根原ら⁹⁾はこの手法を用いて、2017年の東京マラソンにおける訪日外国人の動きを可視化した。ここでは、通信事業会社が訪日外国人向けに提供している無料Wi-Fiサービスで取得されるシステムログと、登録された国籍、性別、年代といった属性情報を用いている。また、藤井ら（2018）は鉄道の駅の改札付近でWi-Fiシステムログデータを計測し、鉄道事業者が保有している改札利用者数と比較することで、実人数を予測するモデルの構築や両者の補完方法の検討などを行っている。

映像解析

映像解析技術の向上に伴い、映像から人数を推定する方法も普及し始めている。現在、映像から個人を識別することは高い精度で行えるようになっているが、混雑した状況では人同士が重複してしまい、人レベルでの識別が困難な状況もある。

そこで、観光統計の収集を目的とする人数計測においては、映像から人レベルで識別した後、人数を数える手法と、映像に映った人の集団からそこに存在する人数を推定する手法の両方

が提案されている（Saleh et al., 2015）。

3-1. ビッグデータを用いる際の注意事項

来訪者数の測定を目的として上記のようなデータを用いる場合、実数をそのまま計測できるケースは少なく、補正や推定が必要となることが多い。その場合、データの代表性については十分に確認したうえで用いる必要がある。たとえば、スマートフォン所有者の年齢による偏りは低減しつつあるが、ソーシャルメディアやWi-Fiの利用については年齢などによる偏りがある。また、データの収集や利用においては個人情報やプライバシーに対する配慮も必要になる。とりわけ、特定の観光地における来訪者の行動などを追跡する場合は、匿名性が担保されなくなる危険性がある。そのため、分析の際には、一定数以上の人の塊をまとめて処理するなど、情報の粒度を粗くするような手続きを踏むことが多い。

また、データや分析方法それぞれに長所と短所が存在するため、複数のアプローチを併用しながら相互に参照しあうことが望ましい。たとえば、国土交通省が実施した「ICTを活用した訪日外国人観光動態調査」¹⁰⁾では、訪日外国人の行動について携帯電話の基地局データ、GPSデータとソーシャルメディアデータの3種類を用いて重層的な分析を行っている。広範囲でのマクロな人の分布には基地局データ、より詳細なマイクロな人の移動についてはGPSデータ、観光地ごとの来訪目的や評価などについてはソーシャルメディアデータを用いて分析している。

ビッグデータ分析と社会調査の併用も大きな意義をもつ。両者の併用による最も大きな利点は、客観的情報と主観的情報の両面からの分析が同時に行えることである。来訪者数などの客観的情報の収集はビッグデータ分析が適しているが、来訪者の動機や満足度などの主観的情報の収集は社会調査が適している。ソーシャルメディアデータからも主観的情報を得ることは可能であ

るが、ソーシャルメディア利用者に対象が限定されたり、知りたい事柄が必ずしも投稿されているとは限らない。仮に、ある観光地に魅力を感じた人のみがソーシャルメディアに投稿をする場合などは、それ以外の人の意見はデータに含まれないことになる。観光地の魅力を高める施策立案には両者の比較が重要になるが、表出されない主観的情報は社会調査からしか得ることができないため、ビッグデータ分析だけでなく社会調査との併用が不可欠である。

4 佐原の大祭におけるデータ収集事例

ここでは、社会調査とWi-Fiシステムログデータを用いて佐原の大祭における観光統計を収集した事例について紹介する。佐原の大祭とは、約300年前から千葉県香取市佐原地区で行われる伝統的な祭礼で、春と秋の年2回行われている。期間中は街の中を10台以上の巨大な山車が曳き廻され、これらはユネスコ無形文化遺産や国の重要無形民俗文化財として登録されている。祭りへの来訪者数は警察から発表されているが、その集計方法は明確に示されておらず、観光施策への活用においては信頼性が不足している。また、来場者の性質についてもデータが収集されていないのが現状である。

そこで、我々は観光政策支援につながる知見を得ることを目的として、2018年の秋(10月12日～14日)に行われた佐原の大祭の来訪者に関する、以下のデータを収集した¹⁾。

- (1) 来訪者に対する社会調査：来訪者の属性や来訪の動機などの測定を目的とする。
 - (2) Wi-Fiシステムログデータ：来訪者数の推定や来訪者の属性などの測定を目的とする。
- 以下ではそれぞれの調査データの詳細を述べる。

4-1. 社会調査

祭りの会場全域で来訪者を対象とした調査を行い、219人から回答を得た。主な調査内容は、

来訪者の属性(性別、年齢、リピーターかどうかなど)、来訪の動機などである。リピーターかどうかについては、「大祭は初めて来訪」、「佐原自体初めて来訪」、「初めてではない」の3項目で測定し、初めてではない人をリピーターとして扱う。来訪の動機については、「知人の紹介」、「過去に来たことがある」、「広告で見た」、「SNSで見た」、「口コミ」、「その他」の項目から複数選択方式で測定した。

4-2. Wi-Fiシステムログデータ

会場付近の5地点34箇所設置されているアクセスポイントから、祭り期間中のWi-Fiシステムログデータを得た。また、ユーザー情報として、年齢や性別や居住地域(市区町村単位)などの属性情報も登録されているので、社会調査の結果と照らし合わせることで両者のサンプルの特徴の違いについても検討可能となる。

4-3. 来訪者数の推定

まずはWi-Fiシステムログデータを用いて来訪者数を推定する。アクセスポイントで検知された人数から推定人数を算出する方法は藤井ら(2018)に従い、以下の式とパラメータを用いる。

$$\text{推定人数} = \text{欠損度} \times \text{Wi-Fi検知人数} / (\text{キャリアシェア率} \times \text{Wi-FiON率})$$

欠損度：2.05(平日), 1.85(祝日)

キャリアシェア率：0.223

Wi-FiON率：0.7

検知された人数とそこから推定された滞在人数は表1のとおりである。

ただし、今回のデータでは同一人物が複数日

表1 Wi-Fiアクセスポイントで検知された人数と推定した人数

	検知人数	推定人数
10月12日	635	8,339
10月13日	1,384	18,176
10月14日	1,372	18,018
合計	3,391	44,533



滞在し、異なる日に検知された場合には重複してカウントされているため、推定人数の合計をそのまま来訪者数の実人数として用いることはできない。言い換えれば、表の44,533人は3日間の延べ人数といえる。

4-4 来訪者の特性

来訪者の特性として、男女比と年代比、来訪者の居住地域、リピーターの割合と来訪の動機について分析した結果を示す。Wi-Fi登録データと社会調査から得られた性別と年代の比率はそれぞれ以下のとおりである。男女比についてはWi-Fiシステムログデータと社会調査では大きな違いは無く、わずかに女性の方が多い結果となっている(表2)。一方で、年代比については大きな違いがあり(図1)、60代以上の割合が社会調査では32%と多いが、Wi-Fiではわずかに3%である。この違いは、年齢が高いほど社会調査に協力的であることと、年齢が低いほどスマートフォンの所持やWi-Fiを利用しやすい

いことの両方が原因となっている可能性がある。この結果は、社会調査とWi-Fiデータの両方のサンプルにバイアスが存在することを意味し、両者を補完的に用いることの有効性を示している。

来訪者の居住地域については、社会調査では都道府県単位で質問した。結果は表3のとおりである。同一県内など、多くの人は近距離から来ていることがわかる。一方、Wi-Fiシステムログデータでも登録されている居住地域が収集できるが、同時に10人以上検知された場合のみ記録されるため、比較的来訪者数の多い近距離の地域のみが記録されている。図2は地図上にプロットした結果である。これは祭りの期間を通じて検知された人数である。表3と図2のいずれの結果も、近距離からの来訪者が多数を占めていることを示している。

次に、リピーターの割合について示す。来訪者がリピーターかどうかはWi-Fiシステムログデータからは得られないため、社会調査のデータのみを示す(表4)。10~20代の来訪者のうち39%

表2 来訪者の男女比

	男	女
社会調査	46%	54%
Wi-Fiシステムログ	49%	51%

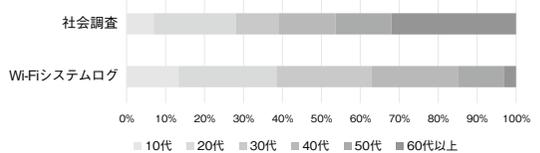


図1 来訪者の年代比

表3 社会調査から得られた来訪者の居住都道府県

都道府県	人数	割合
千葉県	108	58%
東京都	35	19%
茨城県	14	7%
埼玉県	9	5%
神奈川県	7	4%
その他	14	7%

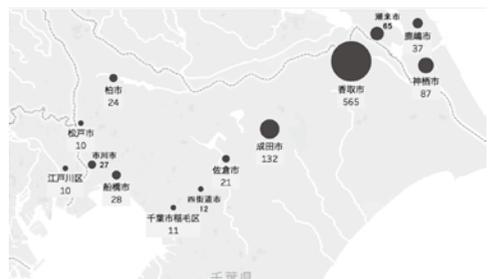


図2 Wi-Fiシステムログデータから得られた来訪者の居住地域(円が大きいほど多くの人 coming していることを意味する)

表4 各年代におけるリピーターの割合

10~20代	30~40代	50~60代以上	全体
39%	48%	53%	48%

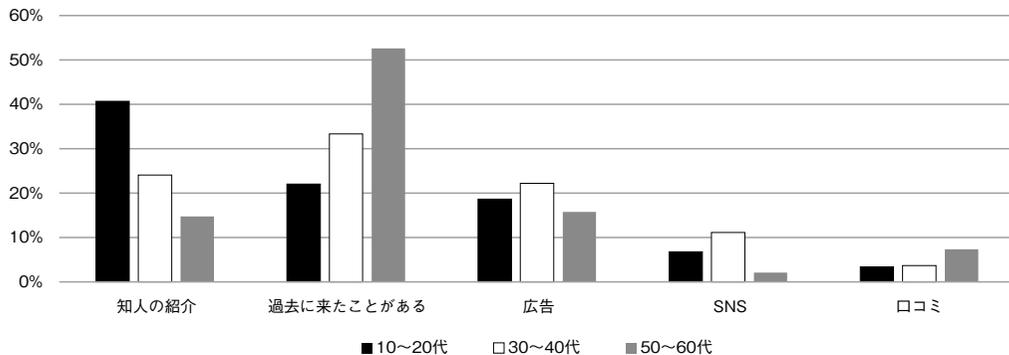


図3 年代ごとの来訪の動機

がリピーターであり、50~60代以上の来訪者では53%がリピーターである。年齢が高くなるほどリピーターの割合が高くなることがわかる。

最後に、来訪の動機として、主にどのようにして祭りのことを知ったのかを社会調査で測定した(図3)。これは主観的情報であり、社会調査でしか測定できないデータである。年齢が高くなるほど、「過去に来たことがある」からと答える人の割合が高く、表4と整合的な傾向が見られる。一方で、「知人の紹介」に関しては、年齢が低いほど多く、若い人はSNSや広告ではなく、知人に紹介されたことをきっかけに来訪する傾向が読み取れる。

以上のように、社会調査とWi-Fiシステムログデータを用いることで、佐原の大祭について以下の観光統計が得られた。3日間の来訪者の延べ人数は約45,000人である。来訪者の男女比は同程度であるが、年齢比については社会調査とWi-Fiシステムログデータとでは大きな違いがあることがわかる。また、来訪者の居住地域は近距離が多いこと、年齢層が高いほどリピーターの割合は高く、年齢層が低く初めて来訪する人は知人の紹介がきっかけである割合が高いことなどが明らかになった。

5 おわりに

本稿では、観光統計の収集におけるビッグデータ分析の有効性と社会調査との併用について概説した。現在では多様な種類のビッグデータが利用可能になっているが、それぞれデータの取得方法や制約も異なっている。また、ビッグデータであっても常に全数調査と見なせるわけではなく、来訪者の推定などにはいくつかの前提が必要となる。

その意味で、現状では唯一無二の万能な方法は存在せず、情報学的アプローチが進んでいる観光統計の分野においても社会調査と併用することが望ましい。特に、佐原の大祭のデータでも見られたように、社会調査とビッグデータではサンプルの年齢構成などに違いがある。観光に関する文脈では、年代によって行動パターンや意識などが異なることも多いため、得られた結果の解釈はサンプルの違いを考慮したうえで行う必要があるだろう。

また、本稿では紹介しなかったが、同一の対象者にビッグデータの提供と社会調査の両方を依頼することで、個人レベルで両者を紐づけることも可能である。これにより、客観的データと主観的データとの因果関係の特定が容易にな



り、社会科学での活用範囲はさらに広まると考えられる(瀧川,2018)。

謝辞

本稿における分析は、曾根原登、本間花、二瓶真友、石井桃香、安河内彩乃、井上わかほ(いずれも津田塾大学総合政策学部)と著者によって実施されたものである。ここに記して感謝したい。

文献

- Abbasi, A., Rashidi, T. H., Maghrebi, M., & Waller, S. T., 2015, "Utilising location based social media in travel survey methods: bringing twitter data into the play", *Proceedings of the 8th ACM SIGSPATIAL international workshop on location-based social networks*.
- Blumenstock, J., Cadamuro, G., & On, R., 2015, "Predicting poverty and wealth from mobile phone metadata", *Science*, 350(6264): 1073-1076.
- Boase, J., & Ling, R., 2013, "Measuring mobile phone use: Self-report versus log data", *Journal of Computer-Mediated Communication*, 18(4):508-519.
- Bohte, W., & Maat, K., 2009, "Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands", *Transportation Research Part C: Emerging Technologies*, 17(3): 285-297.
- Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M., 2016, "The promises of big data and small data for travel behavior (aka human mobility) analysis", *Transportation research part C: emerging technologies*, 68: 285-299.
- Devesa, M., Laguna, M., & Palacios, A., 2010, "The role of motivation in visitor satisfaction: Empirical evidence in rural tourism", *Tourism management*, 31(4): 547-552.
- 藤井秀夫・溝口雄斗・盛江佳史・吉井英樹・曾根原登, 2018, 「改札機カウントデータとWi-Fiデータの統合分析について」『インターネットと運用技術シンポジウム論文集』, 54-61.
- Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L., 2017, "Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States", *Proceedings of the National Academy of Sciences*, 114(50): 13108-13113.
- Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L., 2008, "Understanding individual human mobility patterns", *nature*, 453(7196):779.
- Hasan, S., & Ukkusuri, S. V., 2014, "Urban activity pattern classification using topic models from online geo-location data", *Transportation Research Part C: Emerging Technologies*, 44: 363-381.
- Jurdana, S. D., & Frleta, S. D., 2017, "Satisfaction as a determinant of tourist expenditure", *Current Issues in Tourism*, 20(7): 691-704.
- de Lira, V. M., Macdonald, C., Ounis, I., Perego, R., Renso, C., & Times, V. C., 2019, "Event attendance classification in social media", *Information Processing & Management*, 56(3): 687-703.
- McDowall, S., 2010, "International tourist satisfaction and destination loyalty: Bangkok, Thailand", *Asia Pacific Journal of Tourism Research*, 15(1): 21-42.
- 宮川幸三, 2009, 「我が国の観光統計をめぐる現状と課題——地域観光統計体系の整備に向けて」『産業連関』17(1-2):3-15.
- 野村幸子・岸本達也, 2006, 「GPS・GISを用いた鎌倉市における観光客の歩行行動調査とアクティビティの分析」『総合論文誌』, 4: 72-77
- 岡野雄気・倉田陽平・直井岳人, 2018, 「観光地への愛着に影響を与える滞在中の経験」『観光研究』30(1): 5-18.
- Osorio-Arjona, J., & Garcia-Palomares, J. C., 2019, "Social media and urban mobility: Using twitter to calculate home-work travel matrices", *Cities*, 89: 268-280.
- Phillips, D. L., & Clancy, K. J., 1972, "Some effects of" social desirability" in survey studies", *American Journal of Sociology*, 77(5): 921-940.

- Saleh, S. A. M., Suandi, S. A., & Ibrahim, H., 2015, "Recent survey on crowd density estimation and counting for visual surveillance", *Engineering Applications of Artificial Intelligence*, 41: 103-114.
- Sonehara, N., Suzuki, T., Kodate, A., Wakahara, T., Sakai, Y., Ichifuji, Y., Fujii, H., & Yohii, H., 2019, "Data-Driven Decision-Making in Cyber-Physical Integrated Society", *IEICE Transactions on Information and Systems*, E102-D(9).
- 鈴木春菜, 藤井聡, 2008, 「『消費行動』が『地域愛着』に及ぼす影響に関する研究」『土木学会論文集D』, 64 (2): 190-200.
- 瀧川裕貴, 2018, 「社会学との関係から見た計算社会科学の現状と課題」『理論と方法』 33 (1):132-148.
- Xu, F., & Fox, D., 2014, "Modelling attitudes to nature, tourism and sustainable development in national parks: A survey of visitors in China and the UK", *Tourism Management*, 45: 142-158.
- 山本泰裕・伊藤弘・小野良平・下村彰男, 2006, 「GPSを用いた新宿御苑における利用者の行動パターンに関する研究」『ランドスケープ研究』 69 (5):601-604.
- 弓田恵里香・原田宗彦, 2015, 「スポーツイベント参加者のディステイネーションイメージが評価、満足度、行動意図に及ぼす影響: 参加型スポーツイベントのスポーツツーリストに着目して」『観光研究』 27 (1) :101-113.

注

- 1) 観光立国推進基本計画 (平成 29年度~32年度) <https://www.mlit.go.jp/common/001177992.pdf>
- 2) 国土交通省「宿泊旅行統計調査」<https://www.mlit.go.jp/kankocho/siryou/toukei/shukuhakutoukei.html>
- 3) 国土交通省「旅行・観光消費動向調査」<https://www.mlit.go.jp/kankocho/siryou/toukei/shouhidoukou.html>
- 4) 国土交通省「全国観光入込客統計」<https://www.mlit.go.jp/kankocho/siryou/toukei/irikomi.html>
- 5) https://www.mlit.go.jp/kankocho/page02_000089.html, <https://www.mlit.go.jp/common/001143034.pdf>
- 6) 国土交通省「観光入込客統計に関する共通基準」<https://www.mlit.go.jp/kankocho/siryou/toukei/irikomi.html>
- 7) <https://otonanswer.jp/post/39574/>, <http://www.asahi.com/showbiz/stage/koten/TKY200808260008.html>
- 8) 日本学術会議による提言「社会調査をめぐる環境変化と問題解決に向けて」<http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t248-7.pdf>
- 9) 曾根原登, 国立情報学研究所報道発表「エビデンスに基づく政策・意思決定を支援／多様なデータを構造化して高速処理ソーシャル・ビッグデータ駆動の政策決定支援基盤」http://www.nii.ac.jp/userimg/press_20170314.pdf
- 10) 国土交通省「ICTを活用した訪日外国人観光動態調査」<http://www.mlit.go.jp/kankocho/shisaku/kankochi/gps.html>
- 11) 他に交通手段利用者数やソーシャルメディアなどのデータも取得したが本稿では割愛する。



特集論文

4

日本の地名・人名 データベースから立ち上がる 記憶のコミュニティー

山田太造

東京大学史料編纂所 助教

日本戦国期古記録『上井覚兼日記』のテキストを用い、そこに出現する人物間の関連を検出する手法について述べる。人名間関連検出は、トピックモデルLDA (Latent Dirichlet Allocation) により行う。LDAの結果により、『上井覚兼日記』の観点から検出できた人物グループとその時系列変化、さらには空間的变化について述べる。

1 人と人とのつながりを検出する

あらゆる人と人とのつながりや関係を把握することは困難である。自分を中心として、自分に関係する人との間は明示しやすい。夫婦、親子、兄弟といった血縁関係、同僚や旧友、学友のような関係があろう。ごく親しい人ならそれなりに把握できるだろうが、例えば自分の兄弟姉妹であっても、その友人関係のすべてを把握しているかどうかは甚だ疑問である。スタンレー・ミルグラムは1967年にとても有名かつ重要な実験を行った。スモールワールド現象 (small world phenomenon) の検証実験である (Watts, 2003)。無作為に選んだ2名がつながっている場合、そのつながりの間には平均すると6人の知り合いがいたことから「六次の隔たり (6 Degrees of Separation)」というフレーズでもお馴染みである。これは、「世界は6次元の隔たりに構成されている」ということだけではなく、「知人の知人」は知らない人ばかりかもしれないが、一方で「知人の知人は自分の知人」というネットワークが

存在していることを示唆している。しかも、知人のネットワークは1つではないということも重要である。仕事、血縁、場所などの要因により複数のネットワークが存在する。

近年、インターネットの急速な発展や情報機器の高性能化などにより、ソーシャルネットワークサービス (SNS) も急速に発展している。Twitter, Facebook や Instagram などがある代表例である。SNSの解析を行えば、現時点での人物間の関係を明示的に分析できるかもしれない。では、SNSがない時代の人物間関係を知るためにはどうすべきであろうか。ミルグラムのような実験を再び行うことは不可能である。

我々は、戦国期の薩摩島津家における政治機能を分析するため、経済的・軍事的行動の要因ともいうべき家臣団の形成や支配システムの解明に取り組んでいる。島津家は天正4年 (1576年) の三州 (薩摩・大隅・日向) 統一以降、耳川の合戦や沖田畷の戦いを経て、天正15年 (1587年) の豊臣秀吉による九州征伐までの間に、急速な勢力増強、兵の動員を可能にしている。

例えば、天正12年3月に勃発した沖田畷の戦いでは島津方は3,000人程度の動員だったのに対し、天正14年の九州征伐直前の豊後攻めでは肥後側より30,000人、日向側より10,000人ともいわれる動員を行っている。数年のうちにこれだけの動員を可能にする仕組みとして、家臣団の形成に要因があるかもしれないと考えている。

島津家家臣団を知るには、「人物グループ」と

いう関係性を見出し、どのような場合にどのようなグループが活躍しているのかを検出することがその手がかりになるのではないかと考えた。人物グループを見出す方法として、まず、「あるコンテキストで登場する人物グループは、それと同じようなコンテキストでも登場する」という仮定をおいた。ここでのコンテキストは、外交・合戦・行事などのイベントに該当すると仮定する。あるコンテキストで登場する人物を検出するにはどうすればよいのか。それは歴史資料(史料)に頼るのが良い。前述の島津家の飛躍を知る上で重要な史料の1つとして『上井覚兼日記』がある。この古記録は一般的な日記と同様に、日単位で整序されている。例えば天正14年6月7日条の書き始めは、下記のとおりである。

一、七日、早朝殿中へ出仕申候、白浜次郎左衛門尉殿を以、参上申候由申上候、即御見参被成候、御談合之事、吾々祇候御待居雖被成候、従武庫様急ニ御申之義候間、昨日より御談合にて候、伊地知伯州・吉田作州にて様子承候、…

ここでは何らかの「御談合」に関して白浜次郎左衛門尉、武庫様、伊地知伯州、吉田作州の4名が登場する。ここで「吾々」とあるため、著者である上井覚兼も含めたほうがよいかもしれない。ともかく、これらの人物は「あるコンテキストで登場した」ことを記録することから始める。これは共起関係を調べることと同値であると考ええる。また、このコンテキストを見出すことでどのようなグループであるかを調べる。ここでは「御談合」をコンテキストとすることができるかもしれないが、多くの場合は簡易に検知できない。また、テキストの量が増えれば、人手で整理していくことは困難になる。そこで情報技術、特に機械学習・AI技術の1つであるトピックモデルを用いて、共起関係にもとづく人物グループの自動検出を行う。

2 史料から人物間関連を検出する

2.1 『上井覚兼日記』

本稿では『上井覚兼日記』(図1)を扱う。これは薩摩島津家の家臣であった上井覚兼による日記である。『伊勢守日記』とも呼ばれる。原本は東京大学史料編纂所(以下、史料編纂所)所蔵の重要文化財であり、国宝『島津家文書』に含まれる史料である。また史料編纂所はこの影写本も所蔵している。原本は28冊で構成されている。天正2年8月1日から天正14年10月15日までの約12年間の日記であるが、このうち、天正4年9月7日から天正10年11月3日までの日条は存在しない。

書かれた期間および著者である上井覚兼の立場から、天正2年8月1日から天正4年9月6日までの鹿児島における「奏者時代」、および天正10年11月4日から天正14年10月15日までの「日向国宮崎城主時代」の2つに分けることができる。沖田暁の戦いや岩屋城の戦いに代表される島津家と龍造寺家・大友家との関係や、羽柴秀吉による九州征伐寸前までの島津家の動向など、天正期における島津家や九州の情勢を知る上で重要な史料として位置づけられている。

『上井覚兼日記』は『大日本古記録 上井覚兼日記』(全3冊)として史料編纂所より編纂史料

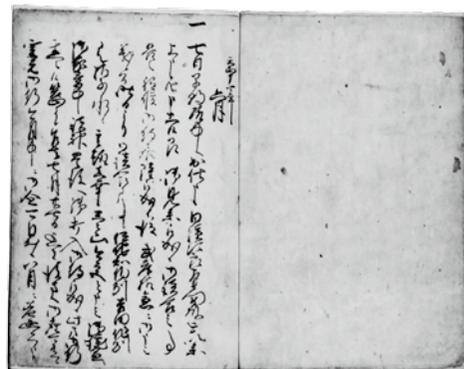


図1 『上井覚兼日記』天正14年6月7日条



集が発刊されており、史料編纂所データベース検索サービス SHIPSDB¹⁾の1つである「古記録フルテキストデータベース」(以下、古記録FT)に、この全テキストが格納されている。『上井覚兼日記』は古記録FTでは「日条」を単位として格納してあることから、これを分析の単位とした。日条は1,777件である。

2.2 『上井覚兼日記』から人名を取り出す

人名データに関しては、JapanKnowledge²⁾・コトバンク³⁾などのオンラインサービスで提供されている「日本人名大辞典」やWikipediaにおける人物の記事のような人名辞典がある。一方で国立歴史民俗博物館「地域蘭学者門人帳人名」⁴⁾・「江戸商人・職人」⁵⁾、国文学研究資料館「歴史人物画像データベース」⁶⁾などがある。これらは、それぞれのデータベースで対象としている史料から人名を網羅的に収集し、人名とそれに関連する史料の典拠などが記されている。しかしながら、このような既存の人物事典のみでは、『上井覚兼日記』に登場する人物すべてを網羅できない。

『大日本古記録』の編纂過程においては、史料に登場する人物にはよく知られている人名を付与しており、古記録FTでもそれが表現されている。これを用いて人名を抽出していく。例えば、先の天正14年6月7日条では、下記の記載がある。

一、七日、早朝殿中へ出仕申候、白浜次郎左衛門尉(重治)殿を以、参上申候由申上候、即御見参被成候、御談合之事、吾々祇候御待居雖被成候、従武庫様(島津義弘)急ニ御申之義候間、昨日より御談合にて候、伊地知伯州(重秀)・吉田作州(清存)にて様子承候、…

この時代では、実名^{いみな}・諱^{なま}で呼びあうのを避ける習慣があるため、仮名^{けみやう}で呼びあうのが通例である。そのため人名を見ても幾人にも該当する

ことがあり得る。編纂過程では特定の人物に比定した結果としてよく知られている名前、例えば括弧を用いて実名を表記している。例えば「白浜次郎左衛門尉(重治)」は「白浜重治」のことを指しており、「伊地知伯州(重秀)」は「伊地知重秀」、「吉田作州」は「吉田清存」であることがわかる。また「武庫様(島津義弘)」の場合は「島津義弘」を示している。

しかしながら、機械がこれを見て判断することができるかどうかの問題である。機械は文章中のどこまでが人名であるかを判断できないためである。そこで、人名記載のルールを作成し、これを機械に学習させ、人名を抽出させた。人名の異なり数は520、述べ4,025だった。

テキストの単位ごと、つまりは日条ごとに人名抽出を行い、そのテキスト内での出現頻度をカウントした。例えば、天正14年6月7日条であれば、白浜重治、伊地知重秀、吉田清存に対して、出現した回数として1を割り当てる。島津義弘はこの日条内に2回の登場であることから2を割り当てる。機械学習、自然言語処理、データ工学などの情報学分野ではこの結果をbag-of-words (BOW) と呼び、テキストを数値データ化する方法の1つとして知られている。

表1は『上井覚兼日記』内に出現する人名とその頻度を示す(上位30件)。この史料の著者である上井覚兼が最も出現頻度が高かった。史料内では「上井覚兼」もしくは「覚兼」ではなく、「拙者」が出現する。上井覚兼が島津家の家老であったことに依ると思われるが、島津4兄弟(島津義

人名	頻度	人名	頻度	人名	頻度
1 上井覚兼	1377	11 肝付兼寛	40	21 島津歳久	27
2 島津義弘	82	12 鎌田兼政	40	22 本田兼親	26
3 伊集院忠棟	74	13 上原尚近	39	23 本田正親	24
4 島津家久	71	14 伊地知重秀	37	24 上井兼成	23
5 島津義久	70	15 上井兼兼	35	25 吉利久金	22
6 島津忠長	59	16 伊集院久宣	35	26 有馬晴信	22
7 柏原有閑	50	17 伊集院久治	34	27 山有信	20
8 本田親貞	45	18 平田光宗	31	28 新納久時	20
9 比志島義基	43	19 平田宗張	30	29 白浜重治	20
10 吉利忠澄	42	20 喜入季久	27	30 平田増宗	19

表1 人名の出現頻度

久, 義弘, 歳久, 家久)のほか, 伊集院忠棟や島津忠長などの島津家の中枢にいた人名の出現頻度が高かった。また, 上井覚兼の家臣だった柏原有閑などの宮崎衆⁷⁾や上井薫兼などの上井覚兼一門衆の頻度も高かった。宮崎衆の出現頻度が高かった理由は, 上井覚兼が日向国宮崎地頭・宮崎城主だったことであろうと思われる。

2.3 地名データを抽出する

人名と同様に、『大日本古記録』では地名にも注釈を付与していることから, 地名の出現箇所を知ることができる。ただし, 中世の地名を現在の地名に比定することは難しいこともある。Google MapsのようなGISシステムが発達している昨今, 地名とともに緯度・経度を付与することが望ましい。そこで, 地名データとして, JapanKnowledge等で提供されている「日本歴史地名体系」, 国立歴史民俗博物館「日本荘園」⁸⁾・「旧高田領取調帳」⁹⁾, 人間文化研究機構「歴史地名データ」¹⁰⁾などを用いながら, 地名に対してそれぞれに緯度・経度を割り当てた。地名の出現回数は述べ1,683だった。

2.4 人物間関係を見出す

人物間の関係性を検出するため, トピックモデルの1つであるLDA (Latent Dirichlet Allocation) (Blei, 2003) を用いた。人名のBOWに対してLDAを適用する。LDAは予め設定したトピック数に応じて, 人名を分類していく。

本研究で用いたLDAのグラフィカルモデルを図2に示す。ここで灰の円は観測変数, 白の円は未知変数を示す。矩形は繰り返しを, 矩形内の右下の数値は繰り返しの回数を示す。wは唯一の観測変数である人名を示す。zはwのトピックを示す。θは日条ごとのトピックの確率分布, φは全人名でのトピックの確率分布を示す。αおよびβはそれぞれ, θおよびφのパラメータを示し, このモデルではハイパーパラメータに相当する。日条数をD, 日条dの人名数をN_d, トピック

数をKとしたとき, θ_dおよびφ_kは

$$\begin{aligned} \theta_d &\sim \text{Dir}(\alpha) \quad (d = 1, \dots, D) \\ \phi_k &\sim \text{Dir}(\beta) \quad (k = 1, \dots, K) \end{aligned} \quad (1)$$

により生成される。ここで, Dir(⋅)はディリクレ分布を示す。トピックz_{d,i}は下記により生成される。

$$z_{d,i} \sim \text{Multi}(\theta_d) \quad (i = 1, \dots, N_d) \quad (2)$$

ここでMulti(⋅)は多項分布を示す。さらに人名w_{d,i}は

$$w_{d,i} \sim \text{Multi}(\phi_{z_{d,i}}) \quad (i = 1, \dots, N_d) \quad (3)$$

により生成される。

LDAのモデル推定では, 崩壊型ギブスサンプリング (Collapsed Gibbs Sampling; CGS) による解法が知られている (Griffiths, 2004)。この手続きを図3に示す。ここではN_{d,i}日条dにおいてトピックiに割り当てられた人名数, Vは人名数を示す。CGSではθおよびφを周辺化(積分消去)することにより直接求めることなく, 次式によりzをサンプリングすることができる。

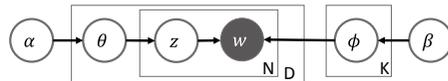


図2 LDAのグラフィカルモデル

1. Initialize α and β
2. Initialize z
3. Set S : the number of sampling
4. for s = 1, ..., S do
5. for d = 1, ..., D do
6. for i = 1, ..., N_d do
7. Sample z_{d,i}
8. Update N_{d,z_{d,i}}
9. end for
10. end for
11. Update α and β
12. end for

図3 Collapsed Gibbs Samplingの手続き



$$z_{d,i} \sim \text{Multi}(p(z_{d,i}|W, Z_{\setminus d,i})) \propto (N_{d,i} + \alpha) \frac{N_{d,i} w_{d,i} + \beta}{N_k + \beta V} \quad (4)$$

LDAのハイパーパラメータである a および β はそれぞれ θ, ϕ のパラメータである。Wallach (2009)により、 a は一様ではなく $\alpha_k \neq \alpha_l, k \neq l$, β は一様である $\beta_1 = \beta_2 = \dots = \beta V$ である場合、LDAのパフォーマンスが改善されることが報告されているため、本研究でも同様の設定とした。さらにCGSの各サンプリングにおいて a および β を更新する。次式に示すように、不動点反復法によりディリクレ分布におけるパラメータを推定することができる。

$$\alpha_k^{new} = \alpha_k \frac{\sum_d \Psi(N_{d,k} + \alpha_k) - D\Psi(\alpha_k)}{\sum_d \Psi(N_d + \sum_{k'} \alpha_{k'}) - D\Psi(\sum_{k'} \alpha_{k'})} \quad (5)$$

$$\beta^{new} = \beta \frac{\sum_k \sum_v \Psi(N_{k,v} + \beta) - D\Psi(\sum_{k'} \alpha_{k'})}{V \sum_k \Psi(N_k + \beta V) - KV\Psi(\beta V)} \quad (6)$$

ここで $\Psi(\cdot)$ はディガンマ関数を示す。

トピックモデルとしてLDA以外にもLSI (Latent Semantic Indexing) (Deerwester, 1990) やpLSI (probabilistic LSI) (Hofmann, 1999) がある。LSIは1史料につき1トピックを仮定するため、多角的な関連性を考慮できない。pLSIはLDAと同様にLSIを拡張し、1史料につき複数トピックを仮定する。しかしながら、潜在トピックの生成確率、つまり式(1)における $p(z|\theta)$ を最尤推定するなどして事前に算出する必要がある。そのため、学習データにはない史料への対応は高コストになってしまう。また、 $p(z|\theta)$ は学習データの量に応じて計算コストが増大してしまうため、アドホックな手法で求めることが多い。これに対しLDAは、 $p(z_k|\theta)$ を確率的に算出する生成モデルである。

3 実験

3.1 LDAを適用する

LDAを『上井覚兼日記』に適用し、人名に関

するトピック、つまりは人物グループを検出する。さらに、時間・空間を軸に人物グループの変化を分析していく。LDAにおけるパラメータとして、トピック数を10とした。図3で示した崩壊型ギブスサンプリングの各繰り返しにおいてパープレキシティ (Perplexity) を計測しており、100回目あたりで急速にその数値が減少し、900回あたりで落ち着いたため、その回数を1,000とした。このサンプリング、またハイパーパラメータである a および β の各要素の初期値は0.1とした。

表2は検出した人物グループと各グループに関係する人名 (出現頻度上位20件分) を示す。例えば、V1は家老や肥後に関わる人名、V2は大隅関係・肝付氏との交渉・抗争に関わる人名、V3は龍造寺氏関係および肥前・筑前情勢に関わる人名、V4は上井覚兼の一門衆・宮崎衆、V5・V8・V9・V10は老中衆・島津一門衆を示すと考えられる。特に、V5は島津歳久関係、V8は当時の島津家の当主である島津義久関係、V9は島津義弘関係、V10は島津家久関係であろうと思われる。V2とV8は一見すると似ているように思えるが、V8は重臣・家老に相当する人物が挙がっているのに対し、V2では大隅・肝付関係者が並んでいるため、上記のように分類できると考えている。

3.2 人物間の関連性を計算する

表3は、島津4兄弟(島津義久・義弘・歳久・家久)と出現パターンが類似する人物の上位15名を示している。この算出方法は下記のとおりである。グループごとの人名とその出現頻度をもとに人物間の距離を算出する。このとき人物-グループのベクトルを作成し、これに次式で示すコサイン類似度を適用する。

$$\text{score}(x, y) = \frac{\sum_i x_i \cdot y_i}{\sqrt{\sum_i x_i^2} \cdot \sqrt{\sum_i y_i^2}} \quad (7)$$

ここで x と y は人物ベクトルを示す。この値は人名の出現パターンに対する類似度に相当す

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	上井覚兼	上井覚兼	上井覚兼	上井覚兼	上井覚兼	上井覚兼	上井覚兼	上井覚兼	上井覚兼	上井覚兼
2	本田親貞	島津義久	吉利忠澄	柏原有閑	島津歳久	吉利忠澄	比志島義基	島津歳久	島津義弘	島津家久
3	平田増宗	伊集院忠棟	秋月種実	鎌田兼政	本田薫親	上井薫兼	伊集院久宣	伊集院久治	上原尚近	島津義弘
4	白浜重治	伊地知重秀	龍造寺政家	上井兼成	阿多忠辰	肝付兼寛	吉利久金	島津忠長	島津忠長	伊集院忠棟
5	伊地知重秀	北郷時久	高城珠長	野村豊綱	喜入季久	鎌田兼政	平田宗張	川上久隅	島津家久	三田井親武
6	八木昌信	本田親貞	大友義統	上井薫兼	土持久綱	伊集院忠棟	新納久時	島津義弘	島津義久	上井神九郎
7	肝付兼寛	喜入季久	島津歳久	桑幡道隆	敷祿頼賀	有馬晴信	大寺大炊助	伊集院忠棟	山忠助	上井秀秋
8	伊集院忠棟	島津以久	稲長辰	田尻鑑種	島津以久	大寺大炊助	川上翌久	山有信	名和顯孝	柏原有閑
9	島津忠長	平田昌宗	有馬晴信	平田宗張	川上久辰	島津義久	山忠助	本田正親	平田歳宗	有川貞真
10	平田光宗	島津義弘	筑紫広門	敷祿頼元	平田光宗	平田光宗	吉利忠澄	本田親正	新納久時	本田正親
11	本田薫親	島津朝久	上原尚近	土持久綱	赤星統家	島津忠長	山有信	合志親重	伊集院久治	税篤和
12	吉田清存	伊地知重隆	并尻祐貞	平田宗応	宮原景種	平田宗衝	上井秀秋	村田経平	伊集院忠棟	島津忠長
13	長谷場純辰	平田宗張	島津義弘	新納忠真	島津家久	敷祿頼賀	吉田清存	猿渡信光	大野忠宗	志岐鎮包
14	税篤和	村田経定	平田宗位	三池鎮実	有馬晴信	鎌田政景	吉利忠富	喜入季久	新納久時	新納忠包
15	龍造寺隆信	東郷重尚	伊地知重則	三田井親武	稲長辰	奈良原延	大寺政安	島津忠良	三原重益	本田親商
16	上原尚近	河野通貞	鎌田政虎	市来忠末	本田盛親	島津家久	小川有季	新納久鶴	北郷忠虎	野村友綱
17	木脇祐充	国分定友	織田信長	桂忠詮	本田親兼	肝付兼篤	市来家守	川田義明	宮内信増	島津久保
18	羽柴秀吉	肝付兼寛	近衛前久	平田宗清	猿渡信光	伊地知重元	柏原有閑	小代親泰	山有信	伊集久治
19	三原重隆	伊地知重興	伊集院忠棟	猿渡信光	鎌田政景	大山綱秀	比志島国貞	川上忠智	島津歳久	北郷久盛
20	島津貴久	本田親治	志岐親重	猿渡信孝	高城珠長	川上忠智	稲長辰	平田光宗	本田正親	吉利久金

表2 LDAにより検出された人物グループ

	島津義久		島津義弘		島津歳久		島津家久	
1	0.93946	本田親正	0.99499	本田親豊	0.99756	本田親兼	0.9113	山善久
2	0.92076	川上久隅	0.87742	深水宗方	0.98432	川上忠智	0.88378	柏原有閑
3	0.89796	島津忠良	0.84678	島津彰久	0.9305	野村重綱	0.79634	田尻鑑種
4	0.89059	大津山家稜	0.80232	島津忠長	0.88588	伊地知重次	0.79343	上井神九郎
5	0.89059	渋谷松右衛門尉	0.78772	高城珠長	0.87977	税篤和	0.79343	野村豊綱
6	0.89059	辺春親行	0.77243	大津山家稜	0.86777	本田薫親	0.79261	上井兼成
7	0.88656	本田正親	0.77243	渋谷松右衛門尉	0.84289	本田親正	0.75798	比志島義興
8	0.884	伊集院久春	0.77243	辺春親行	0.80524	大津山家稜	0.74972	平田宗応
9	0.88164	村田経平	0.76466	村田経平	0.80524	渋谷松右衛門尉	0.74552	鎌田兼政
10	0.88133	山有信	0.76067	山有信	0.80524	辺春親行	0.73595	近衛信尹
11	0.87464	町田忠綱	0.75076	島津義久	0.80485	平田歳宗	0.66292	名和顯孝
12	0.86714	伊集院久治	0.74521	猿渡信光	0.79715	村田経平	0.63726	吉利忠澄
13	0.86572	猿渡信光	0.74236	川上久隅	0.79299	山有信	0.59251	上井薫兼
14	0.82731	宮原景種	0.73279	伊集院久春	0.7809	伊集院久春	0.58867	伊地知重元
15	0.82689	比志島国貞	0.73137	高城珠玄	0.78086	平田光宗	0.55893	有川貞真

表3 島津4兄弟との共起関係

る。表2におけるV9およびV10の結果から、義久と義弘もしくは義弘と家久の出現パターンは類似するかと思われたが、表3の結果から、人名の共起パターンが異なることがわかる。これにより島津4兄弟は独立して行動しているように見受けられる。

島津義弘に近い出現パターンだったのは、当時の島津家当主島津義久だった。義弘は義久に代わって総大将として指揮を執ることが多かったと考えられるが、これが影響するかもしれない。島津家久はその居城が日向国佐土原にあり、上井覚兼の居城は日向国宮崎である。この関係に

よるかもしれないが、家久と類似する人名は表2で示したV4と比較すると、上井覚兼の一門衆・宮崎衆が多く含まれていることがわかる。

図4は著者である上井覚兼を中心とした人物ネットワークを示す。人物間の関連は同様に(7)式により求め、この値が高い上位20名を関連人物としてあり、これを黒円で示している。関連人物においても(7)式の上位20名を算出し、これを灰色円で示している。最も関連度が高かったのは伊集院忠棟であり、本田親貞、税所篤和、平田光宗と続く。彼らは島津家の家老もしくは太刀役であり、主君である島津義久の近くに侍



上井覚兼日記検索システム

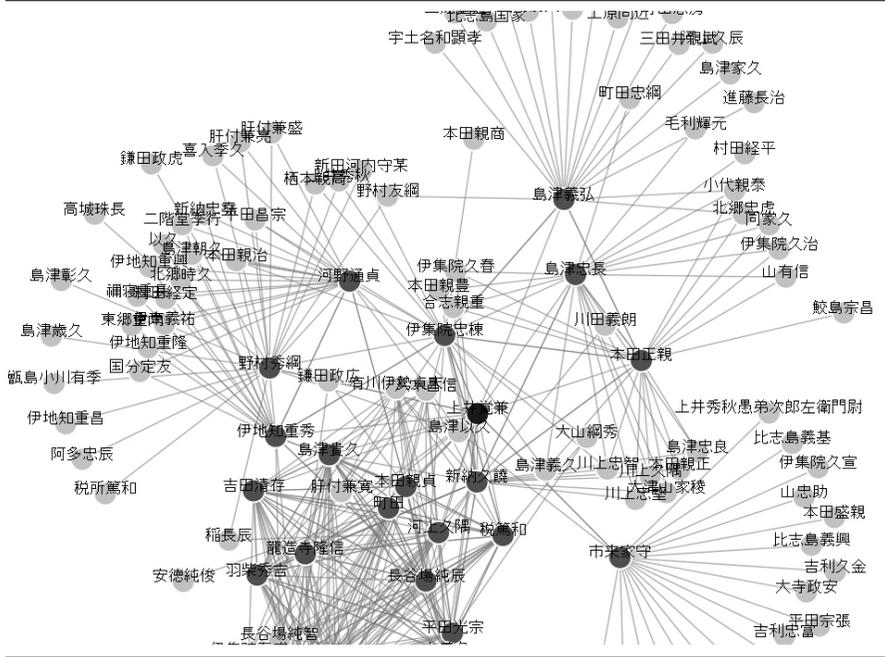


図4 人物ネットワーク

るグループである。表3の結果もここに表示されており、島津4兄弟それぞれの関連する人物を垣間見ることができる。

3.3 人物グループの時系列変化

図5はV9およびV10における時系列変化を示す。この図では、一部の期間について各人物グループに属する人名を列挙した。表2からわかるように、V9は島津義弘、V10は島津家久に関係する人名が属しているが、両トピックとも、義弘・家久が属している。しかしながら、V9にはV1やV2に属する人名が多く、V10にはV4に属する人名が多いことから、これらの特徴は大きく異なる。

図5でもこの傾向が見受けられるが、天正2年から天正4年の期間と天正10年以降の期間とでは両グループともに属する人名が異なることが多かった。例えば、天正3年1月1日－天正3年3月29日の期間においてV9に属する人名は、上井覚兼、上原尚近、平田歳宗、新納教久、島津義弘、

本田親治、伊地知重増、伊地知重秀、伊集院忠棟などである。これに対し、天正14年7月1日－天正14年9月28日の期間では、上井覚兼、島津義久、島津義弘、上原尚近、名和顯孝、島津忠長、伊集院久治、城一要、宮内忠増などである。『上井覚兼日記』の日条において、共起する人名が明らかに異なっていることを示す。これはV10も同様だった。

3.4 地名と人名の関係

図6に、天正10年11月から天正12年4月までの、地名ごとの人物グループの時系列変化を示す。ここでは、この間に本文に出現する地名のうち頻度順に上位6ヶ所(宮崎、紫波州崎、佐土原、鹿兒島、八代、有馬)を選定した。島津家と龍造寺家の間で行われた沖田畷の戦いは天正12年3月24日に勃発した。政情が非常に緊迫した時期である。上井覚兼はすでに宮崎城主であり、かつ島津家の老中役になっている。そのため宮

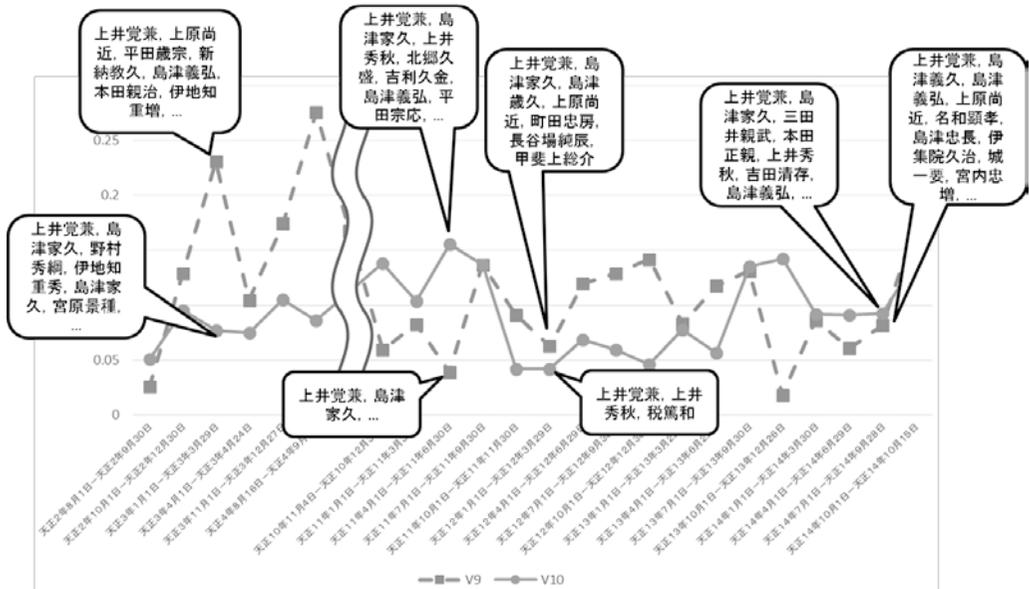


図5 島津義弘と家久の人物グループの時系列変化

崎と鹿児島を頻繁に往来している。

この期間、上井覚兼のお膝元である宮崎で検出された人物グループは頻度順にV1, V4, V2, V6の順だった。V1, V4およびV6は同じ時期に出現していた。一方で、V8はその頻度は低いものの、沖田畷の戦いの直前である天正12年2-3月に突然出現している。史料本文に照らし合わせると、全体的に人物グループの出現頻度が低い期間は宮崎を離れている期間を示し、反対に頻度が高い期間は宮崎在中に相当することがわかった。

紫波州崎は宮崎市折生にある城山を指すと考えられる。宮崎城から見えるほどの近さに位置しており、覚兼の父である薫兼が城主だった。このことから、宮崎衆や上井覚兼の一門衆に関連するグループだと考えているV8の頻度が高い。宮崎在中のときは、紫波州崎まで訪れることが何度もあった。しかしながら、宮崎とは異なる人物グループで構成されていた。とくに天正11年8月付近、および天正12年1月付近で顕著に異なっていることがわかった。これより、宮崎と紫波州崎とは必ずしも同様の政務等を行って

いるわけではないことがわかる。

当時の佐土原は島津家久の居城である。天正11年4月-6月以外の期間では、人物グループ全体としてあまり頻出しなかった。はっきりとしていないが、家久自身が佐土原にいる時期があまりなかったためかもしれない。沖田畷の戦いでは、家久が島津方の実質的な総大将であることも影響すると考えられる。

鹿児島は当主義久の居城であることからさまざまな人物が登場している。最も頻出したグループはV1だった。家久に関連すると思われるV10は、天正11年4-6月の間に頻出し、また天正12年2月にも再び頻出していることがわかる。また、V5は、他のグループに比べると低頻度であるが、宮崎や佐土原などに比べ、比較的頻度が高いのも特徴である。

八代および有馬は上井覚兼の居城の佐土原から見れば、九州の反対側に位置する。こうした地理関係から、これらの地名は天正12年3月の沖田畷の戦い付近の日条で頻出すると予想していたが、実際には天正10-11年の方が圧倒的に頻出していることがわかった。八代においては、

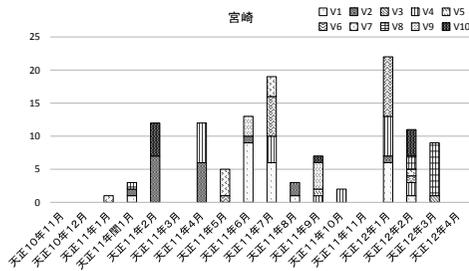


他の地域に比べV10の割合が大きく、天正11年9月にV10が頻出し、天正11年8月にはV1が頻出している。有馬においてもV10の割合は大きい、八代とは異なり、天正11年11-12月に頻出している。また、他の地域に比べV5の割合が大きい。特に、天正12年3月に頻出している。

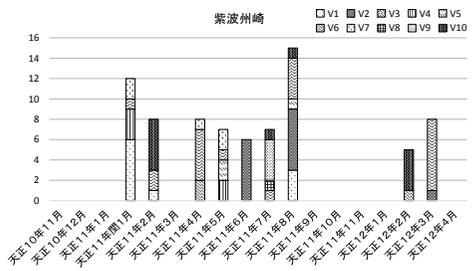
4 おわりに

『上井覚兼日記』の最後の日条は、天正14年10月15日である。秀吉による九州征伐がまさに行

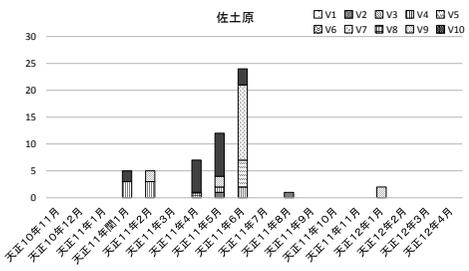
われようとしている時点で終えていることになる。天正14年12月には戸次川の戦いがあり、島津方は秀吉方に一旦の勝利を得たが、天正15年に入ると、秀吉および弟秀長が率いた10万近い軍勢に圧され劣勢となる。根白坂の戦いでの敗戦を経て、この年の4月21日に島津家は秀吉方に降伏を申し出るに至る。それに続く九州国分により、それまでとは異なる人の配置がなされ、既得権益を奪われた旧勢力は国人一揆を起こした。しかしながら、秀吉による徹底した成敗戦略により壊滅していく。九州における近世的秩



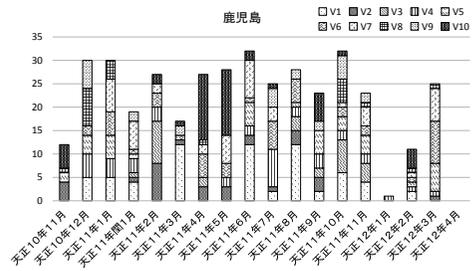
(a) 宮崎



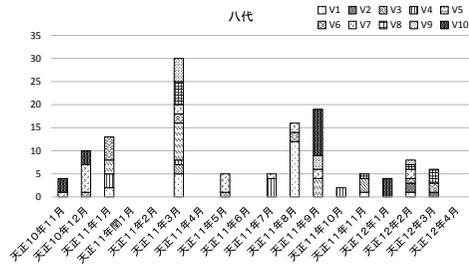
(b) 紫波州崎



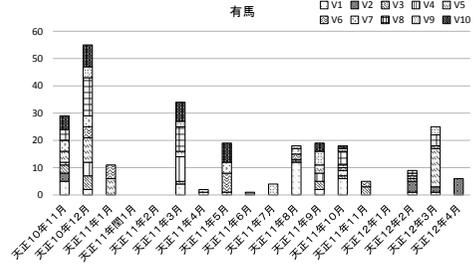
(c) 佐土原



(d) 鹿児島



(e) 八代



(f) 有馬

図6 地名と人物グループの関係

序の始まりを迎えることにより、大大名として残った島津家ですら、旧来の人と土地のあり方を把握することが難しい状態にある。旧来を知る上で『上井覚兼日記』はとても重要な史料である。本稿では『上井覚兼日記』のみを対象としてLDAを用いた分析を行った。それ以外の史料にもあたっていくことで、戦国九州における情勢、さらには島津家の政治的手法を深化し

て解明することができると考えている。

謝辞

本研究の成果の一部はJSPS科研費 18H03576, 26240049, 26730167および東京大学史料編纂所前近代日本史情報国際センターの助成を受けたものによる。

文献

- Watts, D. J., 2003, *Six Degrees: The Science of a Connected Age*, New York:
- W. W. Norton. (辻竜平・友和政樹訳, 2004, 『スモールワールド・ネットワーク — 世界を知るための新科学的思考法』阪急コミュニケーションズ。)
- Blei, D. M., A. Y. Ng, and M. I. Jordan, 2003, “Latent Dirichlet Allocation”, *Journal of Machine Learning Research*, 3(4-5): 993-1022.
- Griffiths, T. L. and M. Steyvers, 2004, “Finding Scientific Topics”, *Proceedings of the National Academy of Sciences*, 101 Supplement 1 (1): 5228-5235.
- Wallach, H. M., D. Mimno and A. McCallum, 2009, “Rethinking LDA: Why Priors Matter”, *Advances in Neural Information Processing Systems*, 23: 1973-1981.
- Deerwester, S., S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, 1990, “Indexing by Latent Semantic Analysis”, *Journal of the American Society of Information Science*, 41(6): 391-407.
- Hofmann, T., 1999, “Probabilistic Latent Semantic Indexing”, *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 50-57.

注

- 1) <https://www.wap.hi.u-tokyo.ac.jp/ships/>
- 2) <https://japanknowledge.com/>
- 3) <https://kotobank.jp/>
- 4) https://www.rekihaku.ac.jp/up-cgi/login.pl?p=param/rung/db_param
- 5) https://www.rekihaku.ac.jp/up-cgi/login.pl?p=param/edos/db_param
- 6) <http://base1.nijl.ac.jp/~rekijin/>
- 7) 『上井覚兼日記』において「宮崎楽」と記述がある。
- 8) https://www.rekihaku.ac.jp/up-cgi/login.pl?p=param/soue/db_param
- 9) https://www.rekihaku.ac.jp/up-cgi/login.pl?p=param/kyud/db_param
- 10) https://www.nihu.jp/ja/publication/source_map



特集論文

5

オープンサイエンスと オープンガバナンスの 倫理的諸問題

近藤康久

総合地球環境学研究所 准教授

1 はじめに オープンサイエンスと社会

オープンサイエンスには、トップダウンとボトムアップの2つの潮流がある(林ほか,2017)。トップダウンの科学技術政策としてのオープンサイエンスは、OECD(2015:9)の「公的資金による研究成果を社会にオープンにすること」を原義とする。ここでいう「オープン」は、Open Definition(2015)において定義されたとおり、「作品に誰もが自由にアクセスでき、作品を自由に再利用、改変、再配布できること」を意味する。

いっぽう、ボトムアップのアクションとしては、〈市民〉、より正確にいえば〈職業研究者ではない人〉が、研究の基礎データを提供したり、観察者・入力者・分析者として研究プロジェクトに参加したりするシチズンサイエンス(市民参加型科学)が国内外で盛り上がりを見せている(Dickinson & Bonney, 2015; Hecker et al., 2018; Ono et al., 2018)。トップダウンとボトムアップのアクションは方向性が異なるように見えるが、学術界(アカデミア)と社会との接合と協働を志向するという点において共通しており、収束の方向に進んでいる。

2015年9月に国連が採択した持続可能な開発目標、いわゆるSDGs(Sustainable Development Goals)には、17個のゴールが掲げられている。そのうちの17番目に「パートナーシップで

目標を達成しよう」とうたわれている。そして2030年までに達成すべき目標(ターゲット)の17.16項には、「知識・専門性・技術・財源の流動化と共有によるグローバル・パートナーシップの強化」が挙げられている(United Nation, 2018)。学術界は「知識・専門性・技術」を担う主要セクターであるから、SDGsの達成に向けて、学術界には社会との連携・連帯が強く求められている現状にあるといえよう。

2 社会課題解決のための研究方法論

社会の課題を解決するための研究の方法論には、長い歴史がある(表1)。その始まりは、グループダイナミクス理論の創始者であるクルト・レヴィン(Kurt Lewin)が、1948年にアクションリサーチ(action research)を提唱したことにさかのぼる(Lewin 1948, 訳1954)。アクションリサーチは、「望ましいと考える社会的状態の実現を目指して研究者と研究対象者とが展開する共同的な研究実践」(矢守, 2010: 1)と定義される。

科学社会学の文脈では、マイケル・ギボンズ(Michael Gibbons)らが、社会に開かれた知識生産の様式(モードII)を提唱した(Gibbons et al. 1994, 訳1997)。モードII科学は、学問体系に貢献することを目的とする従来の科学の知識生産様式(モードI)とは性質が異なる。モードII科学においては、「課題の設定ならびに解決は特

表1 社会課題解決のための研究方法論

アプローチ	特徴	文献
アクションリサーチ Action Research	現実世界の課題解決を志向する研究	Lewin, 1948 矢守, 2010
モードIIサイエンス Mode II Science	学問体系に貢献することを目的とする従来の知識生産のモード(I)とは異なり, 社会に開かれた知識生産のモード	Gibbons et al., 1994 サトウ, 2012
トランス・サイエンス Trans-science	科学のみでは解決できない問題群に対するアプローチ	小林, 2007
シチズンサイエンス Citizen science	市民参加型の共同研究	Dickinson & Bonney, 2015 Hecker et al., 2018 Ono et al., 2018
超学際研究 Transdisciplinary research	現実世界の課題を解決するための, 学術横断的なチームサイエンス	Lang et al., 2011 Mauser et al., 2013 佐藤, 2018

定の学範(ディシプリン)ではなく, 社会の要請によって規定される」(サトウ, 2012: 4)。

また, 科学技術コミュニケーションの文脈では, 2011年の東京電力福島第一原子力発電所事故などを契機として, 科学技術に対する市民の信頼が揺らいだことなどを背景として, 「科学によって問うことはできるが, 科学によって答えることのできない問題群からなる領域」(小林, 2007: 123)であるトランス・サイエンスの意義と役割が見直されるようになった。

近年, 社会課題の解決に資する研究のニーズが高まるなかで, トランスディシプリナリー研究(transdisciplinary research)という言葉をししばしば耳にするようになった。日本ではTD研究, 超学際研究, 超域研究あるいは社会協働研究とも呼ばれるこのアプローチは, 「科学と社会の境界を超えた知識の協働生産プロセス」(佐藤哲, 2018: 4)を本質とする。

TD研究においては, 現実世界の難題(wicked problem)に対処するために, 複数分野の研究者と政府・自治体, 非政府機関, 企業, 地域住民など社会の多様な主体(アクター)がチームを形成して知識経験を持ち寄り, 立場を超えた対話と熟議を通して研究計画の共同立案(co-design), 知識の共同生産(co-production), 成果の共同展開(co-disseminationあるいはco-delivery)を行う(Mauser et al., 2013)。このプロセスにおいては, 研究者が社会の多様な主体に科学

的・専門的知識を一方向的に提供するだけではなく, 社会に備わるさまざまな実践的知識から学びを得て, 互いに学び合う姿勢(mutual learning)が重要とされる(Lang et al., 2012)。

3 シビックテックとオープンガバナンス

また近年, 社会課題解決の文脈において, 「シビックテック」と「オープンガバナンス」という言葉が聞かれるようになった。シビックテックとは「市民主体で自らの望む社会を創りあげるための活動とそのためのテクノロジー」(稲継編, 2018: ii)を意味する。実際には, 情報通信技術(ICT)や社会起業などの知識・技能を持つ市民エンジニアが, 政府・自治体のオープンデータ(オープンガバメントデータ)とICTを活用して, 社会課題の解決に取り組むことを指すことが多い(Knight Foundation, 2013; 松崎, 2017; 稲継編, 2018)。

シビックテックの動きは, ティム・オライリー(Tim O'Reiley)がウェブの概念を行政にも拡張した「Government 2.0」(オライリー, 2009)を提唱したことをきっかけとして米国でおこった。「Government 2.0」構想は, バラク・オバマが2009年1月に米国大統領に就任するにあたり, 透明性(transparency)・参加(participation)・協働(collaboration)の三原則を新たな行動原理として行政官に求めたこと(奥村・米山 2014; 奥



村,2017;宇野他,2017)に呼応するものであった。宇野らが指摘するように、このことは「たんに政府の情報が公開されるだけでなく、利用可能なデータとして提供されることで、市民自らが現状の分析や政策課題の提案をしていくこと」(宇野他,2017:154-155)を意味する。これは、先述した「オープン」の原則に準拠するものであり、「オープンガバメント」と呼びうるものである。

日本では2013年に金沢市にCode for Kanazawaが設立され、シビックテックの実践が始まった(鈴木・福島,2018)。現在、Code for Japanおよび各地域のCode for 団体(ブリゲイド), Civic Tech Forumといったコミュニティの活動や、アーバンデータチャレンジ(UDC)やチャレンジ!オープンガバナンス(COG)といったコンテストを通して、シビックテック実践者の地域のおよび全国的なネットワークが形成されつつある。

COGを主宰する奥村裕一は、「オープンガバメントを行政自身の変革に終わらせず、市民も主役になって課題解決に取り組んでいく社会的仕組みに変えていきたい」(奥村2017:41)という思いを込めて、「市民側から見た社会のガバナンスのオープン化」を「オープンガバナンス」と名付け、「政府視点のオープンガバメント」と区別した(奥村・米山,2014:104)。

市民が主体となって社会課題の解決に取り組むという点において、シビックテックとオープンガバナンスは概念的に共通する。シビックテックがオープンデータとICTの活用を前提とするのに対し、オープンガバナンスはそれらの活用を必ずしも前提としない¹⁾。この概念整理に基づく、シビックテックはオープンガバナンスを実現するための手段のひとつと位置づけることができる。

4 オープンサイエンスとオープンガバナンスの関係性

それでは、このようなオープンサイエンスとオープンガバナンスの時代にあって、社会課題の解決

に資する研究の方法論は、どのように変わるのだろうか。この問いを探究するために、筆者は総合地球環境学研究所(以下、地球研)において「環境社会課題のオープンチームサイエンスにおける情報非対称性の軽減」(通称、オープンチームサイエンスプロジェクト)という共同研究に取り組んでいる。プロジェクトにおける事例・概念の検討(Kondo et al.,2019)を経て、次のことがわかってきた。

まず、オープンサイエンスとオープンガバナンスの目的を比較してみよう(図1)。オープンガバナンスは、社会課題の解決を目的としており、その手段がオープンガバメントデータとICTに基づくシビックテックである。これに対し、政策としてのオープンサイエンスは、日本の第5期科学技術基本計画にうたわれるとおり、オープンリサーチデータをオープンイノベーション、すなわちオープンな科学技術革新につなげることを企図しており、オープンイノベーションの方向性のひとつが社会課題解決である。社会課題の解決に向けて研究者と社会の多様な主体が協働するのがTD研究の要諦であるが、これをオープンリサーチデータと結びつけるためには、市民参加型の共同研究であるシチズンサイエンスと連結する必要がある。TD研究とシチズンサイエンスを連結することにより、社会課題解決型研究と市民参加型研究を融合させ、社会課題解決共創型研究をかたちづけることができる。

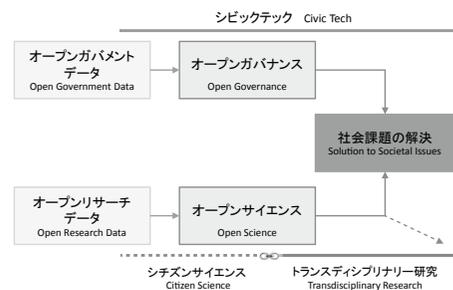


図1 社会課題解決に至るオープンサイエンスとオープンガバナンスの射程(近藤・林,2019:図表3を改変)

次に、シビックテックと、TD研究およびオープンサイエンスの方法論上の異同について検討してみよう(表2)。従来の学術研究(モードI)は研究者が(1)問題を設定し、(2)問題を解くための方法を決め、(3)問題を解き、(4)成果を公表する、というプロセスからなる。これに対し、現在の代表的なTD理論(Mauser et al. 2013)においては、研究者と社会の多様なステークホルダー(利害関係者)が、(1')問題を共同設定(co-design)し、(2')知識を共創し(co-production)、(3')成果を共同発表する(co-dissemination)ことを強調するが、(2)の「問題を解くための方法を決める」工程は明記されておらず、暗黙的に研究者の専決事項とされている。

かたやオープンサイエンスにおいては、TDのような首尾一貫した研究方法論はまだ完成していないが、(3)の問題を解く工程に市民が参加し(シチズンサイエンス)、(4)の成果公表をオープンデータのかたちで行うことが提案されている。

ムキ・ハクレー(Muki Haklay)によれば、シチズンサイエンスは次の4段階に分かれる(Haklay, 2013; 林, 2018)。

- レベル1：市民が〈センサー〉としてデータを収集する「クラウドソーシング」(レベル1)
- レベル2：基礎的なデータ処理を行う「分散型インテリジェンス」
- レベル3：課題設定とデータ収集に加わる「参加型研究」

- レベル4：課題の設定、データの収集・分析を研究者と共同で行う「究極のシチズンサイエンス」

この分類においては、高次のレベルほど市民が研究に深く関与するが、ここでも問題を解くための方法を市民が決めることは想定されておらず、研究者の専決事項であることが暗黙的に了解されているようである。

この点において、シビックテックでは、問題を解くための方法を決める工程も含め、すべてのプロセスを市民が主導し、多様な主体が共創する。この点が、シビックテックと既存の社会課題解決型研究及び市民参加型研究との相違点であり、オープンサイエンスとオープンガバナンスの時代において、社会課題解決共創型研究の方法論をかたちづくる際に留意すべきポイントである。

5 市民と研究者の間の非対称性

オープンガバナンスにおいては、市民と行政、企業、専門家は、理想的には対等な立場で課題解決に参画する。研究者は専門家の一部にすぎない。

しかるに、オープンサイエンスは〈研究〉の一環なので、研究を生業とする職業研究者の地位が先天的に優位になる。社会課題解決型研究においては、取り組む対象が社会の課題であるため、課題については参加者が多かれ少なかれ理解可能である。ただし、問題を解くための方法を理

表2 学術研究のモードIとトランスディシプリナリー(TD)研究、オープンサイエンス、シビックテックのワークフロー比較

従来の学術研究 (モードI)	現在のTD理論 (Mauser et al. 2013)	オープンサイエンス	シビックテック
問題を設定する	Co-design of research agenda	---	市民主導による共創
問題を解くための 方法を決める	---	---	↓
問題を解く	Co-production of knowledge	市民参加	
成果を公表する	Co-dissemination of results	オープンデータ	



解するには、高度な科学的知識を必要とする場合があり、市民と職業研究者の間には知識の非対称性が介在する。

また一般に、職業研究者は高学歴であり、得てして「先生」の敬称を付けて呼ばれる大学教授などの職業に就いているため、シチズンサイエンスの場において、市民参加者との間に社会経済的地位 (socioeconomic status) の非対称性が生じていることもある。

さらにいうと、従来の学術研究 (モードI) においては、このような非対称性が存在することが暗黙の理由付けとなって、市民参加者は正規の研究者と見なされない。そのことは、例えば大学・研究機関が定める謝金支給基準に如実に現れている。筆者の所属する大学共同利用機関法人人間文化研究機構の謝金支給基準では、ワークショップの参加者には「指導・助言謝金」を適用することが認められるが、その額は1日あたり5,000円である。これは、有給の職業研究者に支給することを想定した基準であり、無給の市民やフリーランサーの参加は想定されていない。このように、モードIを前提として構築された大学・研究機関の制度そのものが、市民と職業研究者の間の非対称性を固定化する装置として作用してしまうこともある。

6 欠如モデルと搾取

このような非対称性は、研究者が意識的ないし無意識的に〈上から目線〉で市民に接する構造を醸成するかもしれない。意識的な〈上から目線〉は、直接的な言説として現れるため、受け手が比較的気づきやすいが、無意識的な〈上から目線〉は何気ない言説の中に潜伏するため、受け手にとってもなかなか気づきにくい。

無意識的な〈上から目線〉が潜在する言説の代表例が欠如モデルである。欠如モデルとは、「課題解決に役立つはずの知識・技術を生産しても、それが社会に受け入れられず、活用されないのは、

社会に知識が不足しているからだ」という言説を指す (Sturgis & Allum, 2004; 佐藤, 2018: 1)。欠如モデルは、社会課題解決の文脈にあっては、「地域社会には技術と経験が欠けているから、専門家集団が技術と経験を持ち込めば課題が解決するはずだ」という言説となって発動する。これはオープンサイエンスやTD研究だけでなく、オープンガバナンスとシビックテックにおいても同様である。なお、意識的であるにせよ無意識的であるにせよ、このような〈上から目線〉は専門家本人にとっては自覚しづらいので、注意が必要である。

市民と職業研究者の間に、知識と社会経済的地位などの非対称性があり、欠如モデルが根底にある状況下では、市民からの〈搾取〉の構造が生じやすい (図2)。一般に、市民参加に基づく課題解決を目的とするワークショップは、主催者側の研究者や行政が、市民や学生といった参加者から知識やアイデアを得るという構造になっていることが多い。それは同時に、会場への行き帰りや事前準備・事後フィードバックを含めた一定時間のあいだ参加者を拘束し、知識やアイデアを出させる〈労働〉を課す構造にもなっている。

これに対し主催者は、その場の楽しさや〈やりがい〉、定額の謝金、専門的知識の還元、論文謝辞への記載などのかたちで参加者に報いる。しかし、必ずしも主催者と参加者の間で交換の均衡が取れているとはいえない。交換の不均衡を、弱者の立場にある側が感じるとき、それは〈搾取〉となる。

しかし、普段アカデミアに閉じて暮らしてい

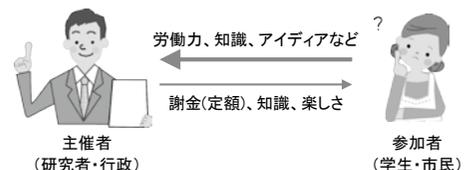


図2 シチズンサイエンスとシビックテックにおける主催者と参加者の非対称性(一例)

る職業研究者は、この〈搾取〉の構造に気づきにくい。なぜなら、先に述べたように、モードIにおける〈研究〉とは自らの知的好奇心に駆動される営みであり、給与・報酬が無くても自発的に行うもの、という先入的価値観があるためである。しかし、この〈無給のボランティア精神〉を、非職業研究者に求めてはいけぬ。なぜなら、職業研究者にとってのインセンティブである論文出版やアカデミアでの就職・昇進は、市民ボランティアにとってのインセンティブとはならないからである。つまり、シチズンサイエンスに参加する市民・学生ボランティアの労力には、参加者側の期待・希望を尊重しつつ、適切に報いる必要がある。

なお、研究者は常に〈搾取〉する側にあるわけではない。研究者が参加者側の立場にある場合には、知識や時間が主催者に〈搾取〉されることもありうる。

7 オープンデータ化の障壁とFAIRデータ原則による解決

オープンガバメントデータとオープンリサーチデータの間にも違いがある。研究データのオープン化には、データ利用者によって引用せずに利用される可能性や先に論文を出版される可能性があり、そこにデータ提供者の抵抗感がある。このことは、ベルモントフォーラムによる国際調査 (Schmidt et al., 2016) のみならず国内調査 (池内, 2018) によっても裏付けられた。データ公開を忌避する傾向は、研究ポストをめぐる熾烈な競争にさらされている若手研究者に特に根強い (Schmidt et al., 2016)。研究キャリアを賭ける成果業績の源泉であるデータを競争相手に奪われたくないという防衛的心理が働くのは、人間として自然なことである。そこで、厳密な意味での「オープン」に代わる現実解として、計算機も人間も発見可能 (findable) ・アクセス可能 (accessible) ・相互運用可能 (interoperable) ・再利用可能 (reusable) であることを条件とするFAIRデータ原則 (Wilkinson

et al., 2016) が提案され、地球惑星科学などの分野においてデファクト・スタンダードになりつつある (Stall et al., 2019)。

8 オープンウォッシング

オープンガバナンスとオープンサイエンスを推進する上でもう一つ留意しなければならないのが「オープンウォッシング」(Openwashing) である。これは、製品や組織が、本当はオープンではないのに、オープンであると騙ることを意味する²⁾。ここでいう「オープン」とは、冒頭に述べたとおり、コンテンツを誰もが自由にアクセス・再利用・改変・再配布できることを意味するが、オープンガバナンスやオープンサイエンスの文脈で社会課題の解決に取り組む現場においては、人権・プライバシーへの配慮やオープン化への抵抗感を克服するために、必ずしも字義どおりの「オープン」では事が進まない場合がある。本来、言葉の意味は時代とともに変わりゆくはずなので、筆者は必ずしも保守的・原理主義的に「オープン」の字義を護るという立場を取らないが、少なくともオープンサイエンスがFAIRデータ原則を現実解とすることを見て取ると、社会課題解決においては〈オープンであること〉よりも〈公平であること〉の方が、優先度が高いように感じる。この点、FAIRデータ原則が字義どおりの公正性 (fairness) の担保となることに期待したい。

9 オープンサイエンスと社会課題解決研究のシナジー

以上の議論をふまえ、地球研のオープンチームサイエンスプロジェクトでは、オープンサイエンスとオープンガバナンス、具体的にはシチズンサイエンスとシビックテックの融合に基づいて社会課題の解決を図る研究方法論〈オープンチームサイエンス〉の構築に取り組んでいる (図3上段)。オープンチームサイエンスは、データだけでなく情報・知識・知恵も含めた〈知〉の総体 (Bellinger

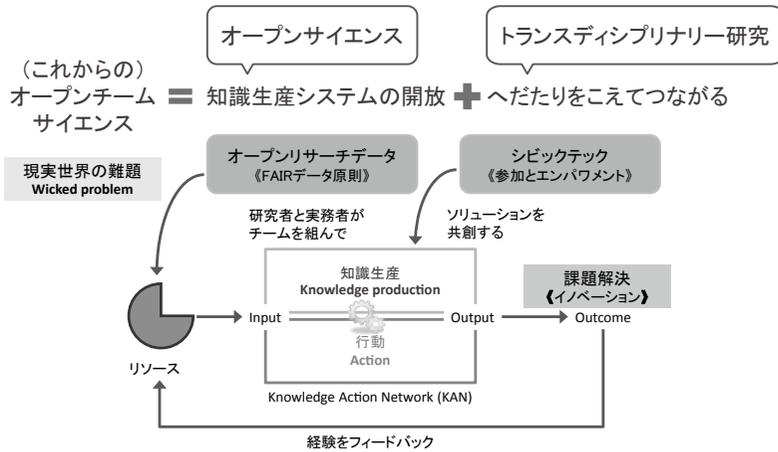


図3 オープンサイエンスとトランスディシプリナリー(TD) 研究の概念的関係性 (近藤・林, 2019: 図表 4)

etal.,2004) を開放するという意味でのオープンサイエンスと、学術界と社会の〈へだたり〉を超えるための方法論としてのTD研究のシナジー(相乗効果)によってかたちづくられる。

〈知〉の総体を開放することは、学術界の知識生産システム自体を開放することを意味する。これがオープンサイエンスの本質である。知識生産システムの開放の一部をオープンリサーチデータが担い、それはFAIRデータ原則にしたがって実現される。他方、シビクテックは〈へだたり〉を超えてつながるための手段と位置づけられる。シビクテックを社会課題解決型研究に取り入れる際には、協働・共創の場において相対的に声の小さい参加者のエンパワメントに特に留意する必要がある。

この図式を現実世界の難題に対処するためのワークフローに当てはめると、以下のようになる(図3下段)。まず、オープンリサーチデータをインプットとして、研究者と社会の多様な主体がチームを組んで知識生産と行動をすすめ、シビクテックを活用してソリューションを共創する。その過程で知識生産と行動のためのネットワーク(Knowledge Action Network ; KAN)が形成される。活動の経験は、次の課題解決の

ためのインプット・リソースとしてフィードバックされる。

10 むすびにかえて

以上をまとめると、オープンサイエンスを社会課題解決に結びつけるためには、市民が主体的に課題を解決するシビクテックの手法を取り入れる必要がある、そのためには職業研究者と市民の間に介在する様々な非対称性に起因して生じる欠如モデルや搾取を排除して、研究者と実務者とが対等な立場で学びあいながら協働を実質化する必要がある、という結論に至る。

筆者らはいま、琵琶湖の水草大量繁殖への対処などの実践を通じて、オープンチームサイエンスの方法論の検証と改良に取り組んでいる。その成果については、機会を改めて報告したい。

謝辞

本稿は既刊論文(近藤・林,2019 ; Kondo et al.,2019) を再構成して新たな考察を追加したものであり、総合地球環境学研究所コアプロジェクトフルリサーチ(14200075)「環境社会課題のオープンチームサイエンスにおける情報非対称

性の軽減」(通称オープンチームサイエンスプロジェクト)による成果の一部である。本稿の内容は、以下に名を記す各氏をはじめとするオープンチームサイエンスプロジェクト及びKYOTO OPEN SCIENCE 勉強会各氏との議論の賜物である。記して感謝申し上げる。

池内有為, 一方井祐子, 牛島健, 宇高寛子, 榎戸輝揚, 大澤剛士, 太田和彦, 大西秀之, 奥田昇, 小野英理, 加納圭, 北本朝展, 熊澤輝一, 佐藤賢一, 下山紗代子, 末次聡子, 関野樹, 中島健一郎, 中塚武, 中西久枝, 中原聖乃, 林和弘, Bianca Vienni Baptista, 宮田晃碩, 村山泰啓 (50音順)

注

1) COG事務局のテクニカルディレクターを務める下山紗代子氏の教示による。

2) <http://openwashing.org> (2019年8月15日アクセス)

文献

池内有為, 2018, 「日本における研究データ公開の状況と推進要因, 阻害要因の分析」『Library and Information Science』79: 21-57.

稲継裕昭編, 2018, 『シビックテック——ICTを使って地域課題を自分たちで解決する』勁草書房.

宇野重規・奥村裕一・犬童周作・関治之・熊谷俊人・久保田后子, 2017, 「オープンガバナンスの時代へ: 現代民主主義にとって大きなチャレンジ」『Voice』471: 154-161.

奥村裕一, 2017, 「市民参加型のオープンガバナンスに向けて」『ガバナンス』200: 41-43.

奥村裕一・米山知宏, 2014, 「オープンガバメントからオープンガバナンスへ: 欧米の動向を踏まえて」『日本情報経営学会誌』34(4): 104-115. (https://doi.org/10.20627/jsim.34.4_104)

オライリー ティム, 2009, 「ガバメント 2.0——政府はプラットフォームになるべきだ」. (<https://jp.techcrunch.com/2009/09/05/20090904gov-20-its-all-about-the-platform/>) (2019年6月24日アクセス)

小林傳司, 2007, 『トランス・サイエンスの時代——科学技術と社会をつなぐ』NTT出版.

近藤康久・林和弘, 2019, 「オープンサイエンスと社会課題解決——マルチステークホルダー・ワークショップによる予察とその後の展開」『STI Horizon』5(1): 35-40. (<https://doi.org/10.15108/stih.00167>)

サトウタツヤ, 2012, 『学融とモード論の心理学——人文社会科学における学問融合をめざして』新曜社.

佐藤 哲, 2018, 「意思決定とアクションを支える科学: 知の共創の仕組み」佐藤哲・菊地直樹編『地域環境学——トランスディシプリナリー・サイエ

ンスへの挑戦』東京大学出版会: 1-15.

鈴木まなみ・福島健一郎, 2018, 「シビックテックって何?」稲継裕昭編『シビックテック ICTを使って地域課題を自分たちで解決する』勁草書房: 1-28.

林和弘, 2018, 「オープンサイエンスの進展とシチズンサイエンスから共創型研究への発展」『学術の動向』23(11): 12-29.

林和弘・宇高寛子・近藤康久, 2017, 「二つのオープンサイエンス, その合流点にある地球研」『地球研ニュース』67: 7-11. (<https://doi.org/10.20568/00001376>)

松崎太亮, 2017, 『シビックテックイノベーション 行動する市民エンジニアが社会を変える』インプレスR&D.

矢守克也, 2010, 『アクションリサーチ——実践する人間科学』新曜社.

Bellinger, Gene, Durval Castro, and Anthony Mills, 2004, "Data, information, knowledge, and wisdom". <http://www.systems-thinking.org/dikw/dikw.htm> (2019年6月25日アクセス)

Dickinson, Janis L. and Rick Bonney, 2015, *Citizen Science: Public Participation in Environmental Research*. Ithaca: Comstock Publishing Associates.

Gibbons, Michael, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott and Martin Trow, 1994, *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: SAGE Publications. (小林信一訳, 1997, 『現代社会と知の創造: モード論とは何か』丸善ライブラリー)



- Haklay, Muki, 2013, "Citizen science and volunteered geographic information: Overview and typology of participation", In *Crowdsourcing geographic knowledge*, edited by Daniel Sui, Sarah Elwood and Michael Goodchild (Eds). Springer: 105-122. (https://doi.org/10.1007/978-94-007-4587-2_7)
- Hecker, Sunanne, Muki Haklay, Anne Bowser, Zen Makuch, Johannes Vogel, and Aletta Bonn, 2018, *Citizen Science: Innovation in Open Science, Society and Policy*. London: UCL Press. (<https://doi.org/10.14324/111.9781787352339>)
- Knight Foundation, 2013, "The Emergence of Civic Tech: Investments in a Growing Field." (https://www.knightfoundation.org/media/uploads/publication_pdfs/knight-civic-tech.pdf) (2019年6月20日アクセス)
- Kondo, Yasuhisa, Akihiro Miyata, Ui Ikeuchi, Satoe Nakahara, Ken'ichiro Nakashima, Hideyuki Onishi, Takeshi Osawa, Kazuhiko Ota, Kenichi Sato, Ken Ushijima, Bianca Vienni Baptista, Terukazu Kumazawa, Kazuhiro Hayashi, Yasuhiro Murayama, Noboru Okuda, and Hisae Nakanishi, 2019, Interlinking open science and community-based participatory research for socio-environmental issues. *Current Opinion in Environmental Sustainability* 39: 1-8. (<https://doi.org/10.1016/j.cosust.2019.07.001>).
- Mausser, Wolfram, Gernot Klepper, Martin Rice, Bettina Susanne Schmalzbauer, Heide Hackmann, Rik Leemans, and Howard Moore, 2013, "Transdisciplinary global change research: the co-creation of knowledge for sustainability". *Current Opinion in Environmental Sustainability* 5 (3-4): 420-431. (<https://doi.org/10.1016/j.cosust.2013.07.001>)
- Lang, Daniel J., Arnim Wiek, Matthias Bergmann, Michael Stauffacher, Pim Martens, Peter Moll, Mark Swilling, and Christopher J. Thomas, 2012, "Transdisciplinary research in sustainability science: practice, principles, and challenges." *Sustainability Science* 7 Supplement 1: 25-43. (<https://doi.org/10.1007/s11625-011-0149-x>)
- Lewin, Kurt, 1948, *Resolving Social Conflicts: Selected Papers on Group Dynamics*. New York: Harper & Brothers. (末永俊郎訳, 1954, 『社会的葛藤の解決: グループ・ダイナミックス論文集』東京創元社.)
- OECD, 2015, Making Open Science a Reality. *OECD Science, Technology and Industry Policy Papers* 25. (<https://doi.org/10.1787/5jrs-2f963zs1-en>)
- Ono, Eiri, Yuko Ikkatai, and Teruaki Enoto, 2018, "Increasing crowd science projects in Japan: Case study of online citizen participation", *International Journal of Institutional Research and Management* 2(1): 19-34.
- Open Definition, 2015, The Open Definition. (<http://opendefinition.org>) (2019年6月20日アクセス)
- Schmidt Birgit, Birgit Gemeinholzer, and Andrew Treloar, 2016, "Open data in global environmental research: The Belmont Forum's Open Data Survey". *PLoS ONE* 11(1): e0146695. (<https://doi.org/10.1371/journal.pone.0146695>)
- Stall, Shelley, Lynn Yarmey, Joel Cutcher-Gershenfeld, Brooks Hanson, Kerstin Lehnert, Brian Nosek, Mark Parsons, Erin Robinson, and Lesley Wyborn, 2019, "Make scientific data FAIR". *Nature* 570: 27-29. (<https://doi.org/10.1038/d41586-019-01720-7>)
- Sturgis, Patrick and Nick Allum, 2004, "Science in society: re-evaluating the deficit model of public attitudes". *Public Understanding of Science* 13(1): 55-74. (<https://doi.org/10.1177/0963662504042690>)
- United Nation, 2018, "SDG Indicators: Metadata repository". (<https://unstats.un.org/sdgs/metadata/?Target=17.16>) (2019年6月24日アクセス)
- Wilkinson, Mark D., Michel Dumontier, Barend Mons, and others, 2016, "The FAIR Guiding Principles for scientific data management and stewardship". *Scientific Data*. 2016/3: 160018. (<https://doi.org/10.1038/sdata.2016.18>)